

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 11:40:42 ; Search time 31.43 Seconds
(without alignments)
1678.653 Million cell updates/sec

Title: US-09-701-572-2
Sequence: 1 MDGTGNNRPPPTSTVRDNP.....KSNTESEIGALSGRTTIR 475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq.032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2555	100.0	475	AAV53985
2	1287.5	50.4	496	AAAB43249
3	1276	49.9	478	ABBS59563
4	1231.5	48.2	1313	ABG13562
5	995	38.9	451	ABBS6028
6	917	33.9	526	ABBS60522
7	900	35.2	499	AAAR78656
8	893.5	35.0	499	AAAR78657
9	888	34.8	511	AAAB43592
10	835	32.7	855	AAAG39932
11	835	32.7	894	AAAG39931

12	835	32.7	898	21	AAAG39930	Arabidopsis thalia
13	254	9.9	74	22	ABG13561	Novel human diagno
14	233	9.1	478	21	AAAG35735	Zea mays protein f
15	229	9.0	479	21	AAAY79678	Drosophila Notchle
16	229	9.0	481	22	ABBS9486	Drosophila melanog
17	228.5	8.9	514	16	AAAR85881	WD-40 domain-contg
18	227.5	8.9	485	22	AAAB68284	Amino acid sequenc
19	226.5	8.9	485	22	AAAB2844	Human protein sequ
20	226.5	8.9	485	22	AAAB68516	Human GTP-binding
21	225.5	8.8	490	21	AAAG3139	Arabidopsis thalia
22	225.5	8.8	490	21	AAAG6519	Arabidopsis thalia
23	222.5	8.7	484	22	AAAB68282	Amino acid sequenc
24	221.5	8.7	439	22	ABBS7838	Drosophila melanog
25	218.5	8.6	471	21	AAAG23141	Arabidopsis thalia
26	218.5	8.6	471	21	AAAG6521	Arabidopsis thalia
27	218.5	8.6	473	21	AAAG3140	Arabidopsis thalia
28	218.5	8.6	473	21	AAAG6520	Arabidopsis thalia
29	215	8.4	514	20	AAAB4268	Candida albicans T
30	212.5	8.3	949	22	ABBS3108	Drosophila melanog
31	209	8.2	486	21	AAAG42985	Arabidopsis thalia
32	209	8.2	499	21	AAAG42984	Arabidopsis thalia
33	209	8.2	700	22	ABBS60376	Drosophila melanog
34	208	8.1	1326	22	ABBS67237	Drosophila melanog
35	208	8.1	1326	22	ABBS7238	Drosophila melanog
36	208	8.1	1326	22	ABBS70051	Drosophila melanog
37	206.5	8.1	339	19	AAAB58888	Human lipid metabo
38	206.5	8.1	339	19	AAAB5957	Human clao-1 prote
39	206	8.1	486	21	AAAG14448	Arabidopsis thalia
40	206	8.1	499	21	AAAG14447	Arabidopsis thalia
41	202.5	7.9	514	22	AAAB5225	Human protein sequ
42	196.5	7.7	261	22	ABBS68576	Drosophila melanog
43	196	7.7	361	22	ABBS68283	Amino acid sequenc
44	196	7.7	680	22	ABBS4206	Drosophila melanog
45	195.5	7.7	347	22	ABBS9902	Drosophila melanog

ALIGNMENTS

RESULT 1	AAV53985	AAV53985 standard; Protein: 475 AA.
ID	AAV53985	
AC	AAV53985	
XX		
DT	13-MAR-2000	(first entry)
XX		
DE	Amino acid sequence of an alfalfa for protein designated ccc52ms.	
XX		
KW	Alfalfa: ccc52ms; WD40 motif; flizy-related; for subfamily; for protein;	
KW	cellular differentiation; endoreplication; cell proliferation;	
KW	plant cell; somatic embryogenesis; endoploidy; reserve tissue;	
KW	in vitro plant regeneration.	
XX		
OS	Medicago sativa.	
XX		
XX		
FT	Key	Location/Qualifiers
FT	Modified-site	43..46
FT	/note= "potential cyclin dependent kinase phosphorylation site"	
FT	Modified-site	128..131
FT	/note= "potential cyclin dependent kinase phosphorylation site"	
FT	Modified-site	144..147
FT	/note= "potential cyclin dependent kinase phosphorylation site"	
XX		
PN	FR2779433-AL.	
XX		
XX	10-DEC-1999.	
PD		
XX		
PF	08-JUN-1998;	98FR-0007174.
XX		

PR 08-JUN-1998; 98FR-0007174.
 XX
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Kondorosi E, Cebolla A, Kondorosi A;
 PI WPI: 2000-089743/08.
 DR N-PSDB; AAZ37024.
 XX
 PT New plant protein, useful for regulating differentiation and
 PT proliferation of plant cells, especially for stimulation of somatic
 PT embryogenesis
 PS
 PS Claim 2; Page 24-25; 36pp; French.
 CC The present sequence represents an alfalfa protein, designated ccc52ms.
 CC The protein contains repeated WD40 (TTP-Asp) motifs, and belongs to
 CC the fzr (fizzy-related) subfamily. The WD40 motifs are situated in
 CC the central portions and the C-terminal portion of the protein. The
 CC protein comprises 7 domains, and the last domain, domain 7, contains
 CC a potential liaison site for cyclins. The N-terminal region of the
 CC protein has a peptide sequence (DRFIPSR) which corresponds to a motif
 CC which is conserved in fzr proteins. The protein intervenes in the
 CC regulation of cellular differentiation, and augments endoreplication.
 CC The polynucleotide sequence and its antisense sequences are used to
 CC regulate the differentiation and proliferation of plant cells and for
 CC stimulating somatic embryogenesis. They are also useful for promoting
 CC endopolydly in plant cells or tissues to increase productivity of
 CC reserve tissues, and to stimulate in vitro regeneration of plants from
 CC callus in culture, by increasing the conversion of embryos.
 CC
 XX
 XX Sequence 475 AA:

Query Match 100.0%; Score 2555; DB 21; Length 475;
 Best local Similarity 100.0%; Pred. No. 8.8e-231;
 Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDGTGNRNPPTSTVRDSSPPPEPSPELSLRHNSRMINSNHTSPRTIYSDRFIPSRAS 60
 DB 1 mdgtgnrnpptstvrdrsspppepspeelslrhnsrminsnhtspirtlysrifpsras 60
 QY 61 KFAFDINTPTREGDSSSAVTLTLRTALFGPDVAGPVTPKTDSPSMTLPNRIERYKT 120
 DB 61 kfaldintptregdrsssaytllrtalfgpdvagvtpktdpsmtlpnrietrykt 120
 QY 121 EFRSGMHSLSPPMDDDFPVGVNHSVPKAPRKVPKRPSPYVLDAPALQODEFYLIVDMSSHN 180
 DB 121 efrsgmhsisppmdddfvgyvnhspvkaprkvprspyvldapalqddfylnlvdmsshn 180
 QY 181 VLAVALGNCVYLMNACSSKVTFLDGLGVDDCVCSYGMARQSTHLAVGNMCKVOIWDAR 240
 DB 181 vlaavlgncvylnnacsskvtkldglgvddcvcsygmargsthlavgnmckvqiwdar 240
 QY 241 CKKISMGBHRLRGALAMSSSLSSGGKDKNIITORDIRTOEDFVSKISGKSEVCGLKW 300
 DB 241 ckkismgbhrlrgalamssslssggkdkniitordirtobdfvskisgksevqglkw 300
 QY 301 SYDNELASGSGNDKLFVWNOHSTQPVLYKCEHRAAVALAMSPRLHGLASGGGTADRC 360
 DB 301 sydnelasgsgndklfvwnohstqpvlykcehraavalaawspmlhglasgggtadrc 360
 QY 361 IRFMMTTNTHSLSCMDTSSQVCNLVMSKNVELVSTHGYSONOITVWYPPMSKLATLTG 420
 DB 361 irfmmttntshslscmdtssqvcnlyvmsknelvsthygsynoitvwyppmsklatltg 420
 QY 421 HTYRLVLYLAISPDGQTIYTAGDETLRFWMNVPSPKSONTESELSGLGRTTR 475
 DB 421 htyrlylylaispdgqtiytagdelrfwmvpspsknteselsglgalttr 475

RESULT 2
 AAB43249

ID AAB43249 standard; Protein; 496 AA.
 XX
 AC AAB43249;
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF3013 polypeptide sequence SEQ ID NO:6026.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnereary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 XX
 XX 05-OCT-2000.
 PD
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 XX 31-MAR-1999; 9905-0127607.
 PR 02-APR-1999; 9905-0127636.
 PR 05-APR-1999; 9905-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX
 XX Shinkets RA, Leach M;
 PI WPI: 2000-602362/57.
 DR N-PSDB; AAC77458.
 DR
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PT
 XX
 PS Claim 11; Page 5209-5211; 5507pp; English.
 PS
 CC AACT4446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnereary;
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypoid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation, to inhibit thrombosis; and as a contraceptive.

Sequence 496 AA;

Query Match 50.4%; Score 1287.5; DB 21; Length 496;
 Best Local Similarity 54.2%; Pred. No. 7.2e-112;
 Matches 269; Conservative 64; Mismatches 110; Indels 53; Gaps 14;

6 NRNPPTSTVDNPPPEPSPESLRHVSRLMINSNHTYSPRTIYSDRFRISRSASKEAL- 64
 17 nentmrvltemrlrltltspasspvs-----spsk--hgdrfflpsraganawsvn 60

65 -FDIN---TPTEGR--DDSSS-----ATTLLRLTALFC---PDVAGPYTPKTKTSPS 107
 61 fhrlnemkspsqnrkakatsdngkldqslaysallknellyagiekvqdpqtdrlrlps.120

108 MTLPRNRNFRKTKTEROSM-----HSLSPMDDFVPGVNHSPYKARXKVRSPRYK 158
 121 -tpekkkltfyslstktrspddgndvpspsnks--qklirspktrkiskipik 177

159 VLDAPALODDFYLNLDVSSHNVLAAGNCVYLMNACSSKVTYKLCDLGVD-DCVCSVGM 217
 178 vldapeidddfylnldvsslnvlsvglctvylwsactsgvtrlcldlsvgeidsvtsygv 237

218 AORGHILAVGTRNKGVOIWDARCKIKSMGHRRLRGALAMSSSLSSGGKDNIIYQD 277
 238 sergnlvavgtlkgfvtlwaagaakklsmlgcharaygalawnaeqslssgsrdtmllgrd 297

278 IRT---QEDFVSKLSGHSSEVGLKWSYDNRELASGDNKLEFVNOHSTOPVLYKCEHT 334
 298 ittpplgee--rtlqghnqevcgklwscdhqlasggnndkllvwnhsislspvglyehl 355

335 AAVKAIAMSPHLGLLASGGGTADRCIRFWNTTNSHLSCHMDGSOVCNLYMSKNVFLV 394
 356 aavkaiawspbhqglasgggtadrcirfwntltgplqicldtsqvcnllawskhanelv 415

395 STHGSONOITWRYPTMSKLTATLGHYRVLYLAISPDGOTIVYAGDETLRFMNVPFS 454
 416 sthysqngllywkyptslqvakltghsryvlylampsdgaivtgadcltrfwnvtsk 475

455 PKSONTESE-IGALSL 469
 476 trstkvkwsesvlnl 491

RESULT 3
 ABB59563
 ID ABB59563 standard; Protein; 478 AA.
 AC ABB59563;
 XX
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 5481.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KM Pharmacological.
 XX
 OS Drosophila melanogaster.
 XX
 PN W0200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL03666.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 5481; 21pp + Sequence Listing; English.
 PS

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 478 AA:

Query Match 49.9%; Score 1276; DB 22; Length 478;
 Best Local Similarity 54.9%; Pred. No. 8.1e-111;
 Matches 260; Conservative 62; Mismatches 114; Indels 38; Gaps 11;

22 PEPSPELRHVSRLMINS--NHTYSPRTIYSDRFRISRS---ASKFALFDIN----- 68
 4 peyekrlkhysparnlfnfessltptsltdrlfporynnwqtnfa--slnksndsp 61

69 -TPTEGRD-----DSSAYTTLRLTALFGPDV-----AGVTPTEKNDPSMTLPNNRIFR 117
 62 qtskkqrdcetardslayscllknellgsaidvktagerenemaytpa---akstlfk 118

118 YKTETRDSM-----HSLSPMDDFVPGVNHSPYKARXKVRSPRYKVLADAPALODDFYLN 172
 119 ygsptkqdygecepslspvsaks--qklirspktrkiskiripkvldeidqldfyln 176

173 LVDMSSHNVLAAGNCVYLMNACSSKVTYKLCDLGVD-DCVCSVGMNORGHILAVGNNG 231
 177 lvdwssqnvlaavglsqvcylwsactsgvtrlcldlspdanltvswnergnvavagvlhng 236

232 KVOIWDARCKIKRSMGHRRLRGALAMSSSLSSGGKDNIIYQDRT-QEDFVSKLSG 290
 237 yvtvwdvaankqklnghaeraygalawnsdlssgsrdtwlqdrtrtpqlserrlag 296

291 HKSEVCGLKWSYDNRELASGDNKLEFVNOHSTQPYLKCERTAAVKAIAWSPHLHGL 350
 297 hrgevcglkwspndqylasggnndrlvwnqhsvnpvqsytehmaavkalawspbhngll 356

351 ASGGGTADRCIRFWNTTNSHLSCHMDGSOVCNLYMSKNVFLVSTHGYSONOITWRYRP 410
 357 asgggtadrcirfwntltgpmgcvdtsqvcnllawskhselvtngysqngllywkyk 416

411 TMSKLTATLGHYRVLYLAISPDGOTIVYAGDETLRFMNVPFSKSONTESEI 464
 417 sltqvaktlghsryvlylalspdgaivtgadcltrfwnvtskarsqkenkv 470

RESULT 4
 ABB13562
 ID ABB13562 standard; Protein; 1313 AA.
 AC ABB13562;
 XX
 XX 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #13553.
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX

FN	MO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001MO-US08631.
XX	
PR	31-MAR-2000; 2000OUS-0540217.
PR	23-AUG-2000; 2000OUS-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dirmamac RT, Liu C, Tang YF;
XX	
DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS77749.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
XX	bioidiversity -

Claim 20; SEQ ID No 43921; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

50	Sequence	1313	AA;
----	----------	------	-----

Query Match	48.28;	Score 1231.5;	DB 22;	Length 1313;
Best Local Similarity	46.38;	Pred. No. 5.5e-106;		
Matches 272;	Conservative 63;	Mismatches 112;	Indels 141;	Gaps 17

QY	6	NNRPETVARNRNSPPPPSPESLHRVSRMINSNHHYSPSPETIISDRPIERSKSKAL-	64
Db	475	nenmpirvtemrtrltltpasspvs-----spsk-ngdrilrpietaganvsvn	518
QY	65	-FDIN-----TPEGR--DDSS-----AVTLLRLALFG--PDVAGPVPPTKTDSPS	107
Db	519	fhriuenekspsgmrkaxdatscngkdglaysallknelldagialekvqpdqdrilrqs	576
QY	108	MTLPRNRNFRKKTROSM-----HSLSPMDDDFVGVGNSPVKARKVPRSPYK	15
Db	579	-tpekvgfflyslstrkspddgnadvspyslspvsnms-qkllrtpkrktrkisklpfk	633
QY	159	VLDAPALODDFYLMVDMSSHNVLAVGLGNCVYLMNACS-----	198
Db	636	vldepelqddfylnlvdwsslnvlsvgltylwsactsqkgermgemgqmgerme	695
QY	199	-----KTRKLCDDIGVD-DCVCSV	215
Db	696	rmgermgemgermgemrdermgemrdwgfetsgkahlclclqvrtrcldsvegdsvstvs	755
QY	216	GMAORG-----THL-----AVGTNKGKVOIMDAARCKKIRSMEG	249

[illegible]

RESULT 5
ABB66028

AC ABB66028;

DT 26-MAR-2002 (first entry)
yy

Drosophila melanogaster polypeptide SEQ ID NO 24876

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

OS *Drosophila melanogaster*.
yy

PN WO200171042-A2.
YY

PD 27-SEP-2001.
YY

PF 23-MAR-2001; 2001WO-US092331.
XX

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150

XX
PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;
vxy

DR WPI; 2001-656860,
DR N-PSDB: ABT10131

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from *Drosophila* and for elucidating cell signalling and cell-cell interactions -

PS Disclosure; SEQ ID NO 24876; 21pp + Sequence Listing; English.

is
The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL016176 ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

Sequence 451 AA;

Query Match	38.98; Score 995; DB 22; Length 451;
-------------	--------------------------------------

PR 14-FEB-1994; 94US-0195730.
 XX (AMGE-) AMGEN INC.
 PA (REGC) UNIV CALIFORNIA.
 XX
 XX
 PI Weinstein J;

DR WPI; 1995-293119/38.
 XX N-PSDB; AAQ96099.

PR Mammalian cell cycle protein p55CDC and DNA sequences encoding it
 PT and methods of modulating cell division using cpds, effecting p55CDC
 CC or associated complexes useful in e.g. chemotherapy
 XX
 XX
 PS Claim 1; Page 54-56; 87pp; English.

CC The p55CDC polypeptide is involved in mammalian cell division. DNA,
 CC proteins and antibodies derived from the polypeptide can be used to
 CC modulate cell cycle activity e.g. in chemotherapy, to inhibit the
 CC growth of tumour cells.
 XX
 XX Sequence 499 AA;

Query Match 35.28; Score 900; DB 16; Length 499;
 Best Local Similarity 39.24; Pred. No. 1.6e-75;
 Matches 195; Conservative 78; Mismatches 175; Indels 50; Gaps 9;

QY 8 NPPSTVIRDNSSPPPEPSRLRVS-----MINSNHTSPSRITSDRF 54
 DB 22 naplarqrakatepapepmraanshsagrtptgpknskvqtlpskp-ggeyrl 80
 QY 55 PPSASAK-----FALFDINTPTEGRDSSAYTTLRTALFEPDVGAPVPEKTDSPSM 109
 DB 81 pprsaagmevasfliskengpedggtprkkehqkawaalnlgfdv----- 125

QY 110 LPRNRIRYKTERROS-----NHSLSPFMDDFVPGVNHSPVKAPRKVPSPYKVLDA 165
 DB 126 -eeakllirsgkpnapegygnrlkylvsqkatpg---ssrkacrylpslptdlidape 181

QY 166 QDDFYRLNLDWMSHNVLAVGLGNCVYLMNACSSKVTKLDL-GVDDCVCSVGAORGT 224
 DB 182 rldylnlvdwssgnvlavaldnsyvlwagsqdlqlqmqepgdyissvawikegnyl 241

QY 225 AVGTNGKVOIMDAARCKIRSMEGHRLRVGALAMSSLSGGRDKNIYQDRTQEDF 284
 DB 242 avgtsaevqlwdvgqgkrlrmtsharvsslswnsyllssgrsghlhhdvraehh 301

QY 285 VSKLSGHKSEVCGLKWSDYNRELASGDNKLFV---NOHSTQPVLYKCYCEHTAAVKA 340
 DB 302 valtsghsgevcglrwapdgrhlsagndnlvnpwspgpegswvplqftqhgavav 361

QY 341 AMSPHLHGLASGGGTADRCIRFWNTTNSHLSCHMTGSOVCNLYWSKNVNLVSTHGS 400
 DB 362 aecpwsnllatgsgtsdhlrlmncsacslavdshvscslwspshykelisghfa 421

QY 401 QNOIIVWRPYTMSKLTATLTGTYRVLYLAISPQGTIVTGAGDETLRFNNVNF---PSKS 457
 DB 422 gnglviwkyptmakvaelkgharvlsltmspdgatvasaadeltlrwrfeldpallr 481

QY 458 QNTESEIGALSIGRTTIR 475
 DB 482 erekastskssllhglr 499

RESULT 8
 AAR78657
 ID AAR78657 standard; Protein; 499 AA.
 XX AAR78657;
 AC
 XX
 DT 23-JAN-1996 (first entry)
 XX

DE Human cell cycle protein p55CDC.

XX Cell cycle; protein; p55CDC; antibody; chemotherapy; modulation;
 KW inhibition; growth.
 XX
 XX Homo sapiens.

XX WO9521917-A1.

XX 17-AUG-1995.

XX 13-FEB-1995; 95WO-US01806.

XX 14-FEB-1994; 94US-0195730.

XX (AMGE-) AMGEN INC.
 PA (REGC) UNIV CALIFORNIA.
 XX
 XX
 PI Weinstein J;

DR WPI; 1995-293119/38.
 DR N-PSDB; AAQ96100.

PT Mammalian cell cycle protein p55CDC and DNA sequences encoding it
 PT and methods of modulating cell division using cpds, effecting p55CDC
 CC or associated complexes useful in e.g. chemotherapy
 XX
 XX
 PS Claim 1; Page 58-60; 87pp; English.

CC The p55CDC polypeptide is involved in mammalian cell division. DNA,
 CC proteins and antibodies derived from the polypeptide can be used to
 CC modulate cell cycle activity e.g. in chemotherapy, to inhibit the
 CC growth of tumour cells.
 XX
 XX Sequence 499 AA;

Query Match 35.08; Score 893.5; DB 16; Length 499;
 Best Local Similarity 38.58; Pred. No. 6.5e-75;
 Matches 187; Conservative 81; Mismatches 165; Indels 53; Gaps 8;

QY 8 NPPSTVIRDNSSPPPEPSRLRVS-----MINSNHTSPSRITSDRF 54
 DB 22 naplarqrakatepapepmraanshsagrtptgpknskvqtlpskp-ggeyrl 80

QY 55 PPSASAKFALDINTPTEGRDSSAYTTLRTALFEPDVGAPVPEKTDSPSKTLPNRN 114
 DB 81 pprsaagmevasfliskengsemsq-----tpkkehqkawaalnlgfdv 121

QY 115 -----IFRYKTERROS-----NHSLSPFMDDFVPGVNHSPVKAPRKVPSPYKVLDA 162
 DB 122 gfdveakllirsgkpnapegygnrlkylvsqkatpg---ssrkacrylpslptdlidape 178

QY 163 PALDDDFYRLNLDWMSHNVLAVGLGNCVYLMNACSSKVTKLDL-GVDDCVCSVGAORGT 221
 DB 179 pelrldylnlvdwssgnvlavaldnsyvlwagsqdlqlqmqepgdyissvawikeg 238

QY 222 THLAVGTNGKVOIMDAARCKIRSMEGHRLRVGALAMSSLSGGRDKNIYQDRTQEDF 281
 DB 239 nylavgsaevqlwdvgqgkrlrmtsharvsslswnsyllssgrsghlhhdvraehh 298

QY 282 EDEVSKLSGHKSEVCGLKWSDYNRELASGDNKLFV---NOHSTQPVLYKCYCEHTAAV 337
 DB 299 ehvatsghsgevcglrwapdgrhlsagndnlvnpwspgpegswvplqftqhgavav 358

QY 338 KATAMSPHLHGLASGGGTADRCIRFWNTTNSHLSCHMTGSOVCNLYWSKNVNLVSTH 397
 DB 359 kavawcpwsnllatgsgtsdhlrlmncsacslavdshvscslwspshykelisgh 418

QY 398 GYSQNOIIVWRPYTMSKLTATLTGTYRVLYLAISPQGTIVTGAGDETLRFNNVNF---PSKS 457
 DB 419 gfaqnglviwkyptmakvaelkgharvlsltmspdgatvasaadeltlrwrfeldpallr 478

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145226.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 32.7%; Score 835; DB 21; Length 855;
Best Local Similarity 40.7%; Pred. No. 4,5e-69;
Matches 171; Conservative 68; Mismatches 89; Indels 92; Gaps 7;

OY 51 DFEISRSASKF--ALPFIPTPTBGRDDSSAYTTLTALFGPDVAPVPEKTDSPSM 108
Db 249 dffprsamfdyahfalteerkykqdsat-----tserl----- 284
OY 109 TLPNRIERYKTKETQSMHSLSPFMDDFVPGVNHSPYKARPKYRSPFYVLDAAPALODD 168
Db 285 -----ldapdlvdd 293
OY 169 FYLNIVDWSHNVAVGAGNCVYLMNACSSKYTKICDLDVD-DCVCSYGAORCTHLAVG 227
Db 294 fylvnlidwgsanvialaithcvylwdaatgstselvtldeekpyvtslnwagdgrrhavg 353
OY 228 TNNKGVQIIMDARCKIKISM--GHRIRYGLAWSSSLISGGRKNITQRIRIQEDPVS 286
Db 354 lnneevqldwsasnmqrltlkygnsrvgslawmhlltvgmgllinnvtrispie 413
OY 287 KLSGKSEVCGIKMSYDREIELASGNDNKLFWNQ-----HSTQPVLEKYEHTAAVKAI 340
Db 414 tyrghtgevcgikwsgsgqqlasgndhvnhlwdsavassnstcgwlhrleehsarkal 473
OY 341 AMSPHLHGLLASGGGTADRCIRFWNTTNSHLSCMDTGSQVCNLVWSKNVELVSTHGS 400
Db 474 awcpfganllatgsggdrtlkfwthtgaclnsvdltsgycslilwsknerelissngft 533
OY 401 ONOIIVMRYPMTSKRLATLGHYRYLYLAISPDQOTYTGAGDFT-----LRFWVVF 453
Db 534 gngclwkyrpsmvmaeltgstrvlymagspdcvtvaagdelclspndkalsfwllnp 593

RESULT 11
AAG39931
ID AAG39931 standard; Protein: 894 AA.
AC AAG39931;
DF 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 49477.
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.

XX EP1033405-A2.
PN 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PD 06-SEP-2000.
XX 23-MAR-1999; 99US-0123548.
PE 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0126264.
PR 01-MAR-1999; 99US-0126265.
PR 01-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134318.
PR 14-MAY-1999; 99US-0134319.
PR 14-MAY-1999; 99US-0134320.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR	01-JUL-1999	99S05-0141842
PR	01-JUL-1999	99S05-0141844
PR	02-JUL-1999	99S05-0142055
PR	06-JUL-1999	99S05-0142390
PR	08-JUL-1999	99S05-0142803
PR	09-JUL-1999	99S05-0142820
PR	12-JUL-1999	99S05-0142977
PR	13-JUL-1999	99S05-0143524
PR	14-JUL-1999	99S05-0143624
PR	15-JUL-1999	99S05-0144005
PR	16-JUL-1999	99S05-0144085
PR	16-JUL-1999	99S05-0144086
PR	19-JUL-1999	99S05-0144321
PR	19-JUL-1999	99S05-0144325
PR	19-JUL-1999	99S05-0144332
PR	19-JUL-1999	99S05-0144333
PR	19-JUL-1999	99S05-0144334
PR	19-JUL-1999	99S05-0144335
PR	19-JUL-1999	99S05-0144352
PR	20-JUL-1999	99S05-0144632
PR	20-JUL-1999	99S05-0144684
PR	21-JUL-1999	99S05-0144814
PR	21-JUL-1999	99S05-0145086
PR	21-JUL-1999	99S05-0145085
PR	22-JUL-1999	99S05-0145085
PR	22-JUL-1999	99S05-0145087
PR	22-JUL-1999	99S05-0145092
PR	22-JUL-1999	99S05-0145158
PR	23-JUL-1999	99S05-0145145
PR	23-JUL-1999	99S05-0145148
PR	23-JUL-1999	99S05-0145524
PR	26-JUL-1999	99S05-0145572
PR	27-JUL-1999	99S05-0145513
PR	27-JUL-1999	99S05-0145518
PR	27-JUL-1999	99S05-0145519
PR	28-JUL-1999	99S05-0145551
PR	02-AUG-1999	99S05-0146386
PR	02-AUG-1999	99S05-0146388
PR	02-AUG-1999	99S05-0146389
PR	03-AUG-1999	99S05-0147038
PR	04-AUG-1999	99S05-0147204
PR	04-AUG-1999	99S05-0147302
PR	05-AUG-1999	99S05-0147160
PR	05-AUG-1999	99S05-0147303
PR	06-AUG-1999	99S05-0147303
PR	09-AUG-1999	99S05-0147493
PR	09-AUG-1999	99S05-0147935
PR	10-AUG-1999	99S05-0148171
PR	11-AUG-1999	99S05-0148341
PR	12-AUG-1999	99S05-0148341
PR	13-AUG-1999	99S05-0148565
PR	16-AUG-1999	99S05-0149368
PR	17-AUG-1999	99S05-0149175
PR	18-AUG-1999	99S05-0149426
PR	20-AUG-1999	99S05-0149722
PR	20-AUG-1999	99S05-0149723
PR	20-AUG-1999	99S05-0149929
PR	23-AUG-1999	99S05-0149902
PR	23-AUG-1999	99S05-0149930
PR	25-AUG-1999	99S05-0150566
PR	26-AUG-1999	99S05-0150884
PR	27-AUG-1999	99S05-0151065
PR	27-AUG-1999	99S05-0151080
PR	30-AUG-1999	99S05-0151308
PR	31-AUG-1999	99S05-0151343
PR	01-SEP-1999	99S05-0151533
PR	07-SEP-1999	99S05-0152363
PR	10-SEP-1999	99S05-0153070
PR	13-SEP-1999	99S05-0153758
PR	15-SEP-1999	99S05-0154018

PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154779
PR	22-SEP-1999	9905-0155139
PR	23-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0155659
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	13-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159331
PR	14-OCT-1999	9905-0159637
PR	14-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160989
PR	25-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161992
PR	28-OCT-1999	9905-0161993
PR	29-OCT-1999	9905-0162142

Query Match	32.7%;	Score 835;	DB 21;	Length 894;
Best Local Similarity	40.7%;	Pred. No. 4.8e-69;		
Matches 171;	Conservative 68;	Mismatches 89;	Indels 92;	Gaps 7;

[illegible]

RESULT 12
AAG3930
ID AAG3930 standard; Protein: 898 AA.
XX AAG3930;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 49476.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EPI03405-A2.
XX
XX
XX 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130444.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134376.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140033.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR	20-AUG-1999;	9905-0149722.
PR	20-AUG-1999;	9905-0149723.
PR	20-AUG-1999;	9905-0149829.
PR	23-AUG-1999;	9905-0149902.
PR	23-AUG-1999;	9905-0149930.
PR	25-AUG-1999;	9905-0150566.
PR	26-AUG-1999;	9905-0150884.
PR	27-AUG-1999;	9905-0151065.
PR	27-AUG-1999;	9905-0151066.
PR	27-AUG-1999;	9905-0151080.
PR	30-AUG-1999;	9905-0151303.
PR	31-AUG-1999;	9905-0151438.
PR	01-SEP-1999;	9905-0151930.
PR	07-SEP-1999;	9905-0152363.
PR	10-SEP-1999;	9905-0153070.
PR	13-SEP-1999;	9905-0153758.
PR	15-SEP-1999;	9905-0154018.
PR	16-SEP-1999;	9905-0154039.
PR	20-SEP-1999;	9905-0154779.
PR	22-SEP-1999;	9905-0155139.
PR	23-SEP-1999;	9905-0155486.
PR	24-SEP-1999;	9905-0155659.
PR	28-SEP-1999;	9905-0156458.
PR	29-SEP-1999;	9905-0156596.
PR	04-OCT-1999;	9905-0157117.
PR	05-OCT-1999;	9905-0157753.
PR	06-OCT-1999;	9905-0157865.
PR	07-OCT-1999;	9905-0158029.
PR	08-OCT-1999;	9905-0158232.
PR	12-OCT-1999;	9905-0158369.
PR	13-OCT-1999;	9905-0159293.
PR	13-OCT-1999;	9905-0159294.
PR	14-OCT-1999;	9905-0159329.
PR	14-OCT-1999;	9905-0159330.
PR	14-OCT-1999;	9905-0159331.
PR	14-OCT-1999;	9905-0159637.
PR	14-OCT-1999;	9905-0159638.
PR	18-OCT-1999;	9905-0159584.
PR	21-OCT-1999;	9905-0160741.
PR	21-OCT-1999;	9905-0160767.
PR	21-OCT-1999;	9905-0160768.
PR	21-OCT-1999;	9905-0160770.
PR	21-OCT-1999;	9905-0160814.
PR	21-OCT-1999;	9905-0160815.
PR	22-OCT-1999;	9905-0160980.
PR	22-OCT-1999;	9905-0160981.
PR	22-OCT-1999;	9905-0160989.
PR	25-OCT-1999;	9905-0161404.
PR	25-OCT-1999;	9905-0161405.
PR	25-OCT-1999;	9905-0161406.
PR	26-OCT-1999;	9905-0161359.
PR	26-OCT-1999;	9905-0161360.
PR	26-OCT-1999;	9905-0161361.
PR	28-OCT-1999;	9905-0161920.
PR	28-OCT-1999;	9905-0161992.
PR	28-OCT-1999;	9905-0161993.
PR	29-OCT-1999;	9905-0162142.

	Query Match	32.7%	Score 835;	DB 21;	Length 898;
	Best Local Similarity	40.7%	Pred. No. 4.8e-69;		
	Matches 171; Conservative	68;	Mismatches 89;	Indels 92;	Gaps 7;
OY	51 DRFIPSRSAKFE--ALFPINTPTGCRDDSSAYTLLRTALRGDPVAGVPPEKDPSPM	108			
	: : :				
Db	292 drfiprsamafdyahafalteerkykdsat-----tsert----	327			
OY	109 TLPRNRIRKYETETQSHSLSPFFMDDFGVGVNHSPVKARKKVRSPLYKLDAFDALQDD	168			
	:				
Db	328 -----Idappivdd	336			
OY	169 FYLVLDWSSHNVLAVLGNGCYLLNACCSKVTKLCIDGLVD-DVCYSVGMAQRGHILANG	227			
	: : : : : : :				

Db 337 fymllldwgsanvjalaladhvylwdastgststslvlldeekgpyvtsinnapgrinhavg 396

Qy 228 TNNKGVQIWDARCKKTRISME-GHRLRVGALAWSSLSLSCGRDKNIYQDIRTOEDFVS 286

Db 397 lnnsevgllwdasannrqrllrtllgqgsqsvslawmnhlltggmglilindvrlspive 456

Qy 287 KLSGKSEVCEILKMSYDRELASGDNPNKLFVWMO-----HSTQPYLKTCERTAAVKAI 340

Db 457 tyrlhtgevcgllkxsgsgqqlaasgndnvnvhlidrstvaassnltqvlhrlleethsavkal 516

Qy 341 AWSPEHLGLLASGGCTADRCIRFNNNTTNSHLSMDTGSOVCNLWMSKNVNELVSTHGIS 400

Db 517 awcfqganllalatggsgggdrllkfmnthtgacloinsvdgsvcsllvskneretllshgft 576

Qy 401 ONOIIIVWVPTMSKLTATLTGGTTVVLEKIASPDQOTIVTGAGDEP-----LRFNWVP 453

Db 577 gnglltllkxyspmvkaeltgltstsvllymagsgpdcvraasaagdeclspndkalsfwllnp 636

RESULT 13
ABG13561
ID ABG13561 standard; Protein; 74 AA.
XX
AC ABG13561;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #13552.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
PI
DR WPI; 2001-639362/73.
DR N-PSDB; AAS77748.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 43920; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and


```

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157153.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158366.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

```

```

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 9.18; Score 233; DB 21; Length 478;
Best Local Similarity 20.98; Pred. No. 5,4e-13;
Matches 113; Conservative 76; Mismatches 161; Indels 190; Gaps 22;

```

```

QY 13 STVRDNPPEPPEPESTRHVSRMINSNHYTSPSPRIYSDRIPR----- 57
DB 2 ataaadgppve--pslkkis-----lrktsl--dlfaphaflfpdaesklrt 50
QY 58 ----SASKFALFDINTPEGRDSSSAVYTLRLTALGPDVAGVPEPEKTDSPMLPNR 113
DB 51 gckvaeygavkdl-tpeggr-----ggqgkypappst--alaipg- 89
QY 114 NIFRYKTETRQSMHSLSPFMDDFVPCVNHSPVAPKPKVPRSPK-----VIDAPALQ 166
DB 90 ----tqtkdang-----egsrnaivpapiimlpkapestipgkntslspgss 133
QY 167 DDFYLNLY-----DMSHNYLAVGLG-----NCVY 191
DB 134 drfsalmeriparwprptchapwkyrvlsghlgwvrsiafdpanewltsadrtlk 193
QY 192 LMNACSSKVTYLCGLDGDVCVSGMAQRGTHLAVGTNNKQVQIWDARCKKIRSMEGHR 251
DB 194 lwdlaagtl-klitghleglrglavagrhylfsagddkqvkwdlegkvlrshyhl 252
QY 252 LRVGALAW--SSSLSSGRDKNIYORDIRFOEDFVSKLSGHSEVCGLKMSDNRPLAS 309
DB 253 sgvyclalhpitdlilggrdsvcrvwdiltka-hvsaletghntvcsvfarptqpyvt 311
QY 310 GGNDNKLFVWQNSTQPYLKCETHAIVKALWASP-----HLH 347
DB 312 gshdtikfwdlvagrncclthhkvramalhpkekafasasadvkfnlpkgeflh 371
QY 348 -----GILASGCGRADRCIRFMNTTSHLSCHMTGSGQVCNLWMSKN 389
DB 372 nmIsqgkllmsmavnedgylatcgdnsg--lwfwd-----wk-- 407
QY 390 VNELVSTHGYSQNOIIVWRYPTMSKSLATLNGHTYRVLYLAISPDGOTIVGADETLRFW 449
DB 408 ----sghnfgdqdtlv-qpgslesaciyalsydv-----sgsrlyvscadcktlkmw 454

RESULT 15
AAV79678
ID AAV79678 standard; Protein; 479 AA.
XX
AC AAV79678;
XX
DT 29-AUG-2000 (first entry)
XX
DE Drosophila Notchless protein.
XX
KW Notchless; Nle gene; Notch; signalling; neurodegenerative disease;
KM cancer; diagnosis; cytostatic; neuroprotective; therapy.
XX
OS Drosophila melanogaster.
XX
FH Key Location/Qualifiers
FT Domain 27..105
FT /note= "Nle domain"
FT Misc-difference 67
FT /note= "sequence deduced from nucleotide sequence
FT has an additional Lys residue between
FT Lys-67 and Ser-68"
FT Misc-difference 282
FT /note= "encoded by CAA"
PN WO200026364-A1.

```

Job time: 339 sec

XX 11-MAY-2000.
 PD 03-NOV-1999; 99MO-1B01891.
 PF 03-NOV-1998; 98GB-0024045.
 XX (EURO-) EURO MOLECULAR BIOLOGY LAB.
 PA Cohen S, Boumeester A, Royet J;
 PI WPI: 2000-365613/31.
 DR N-PSDB; AAA27739.
 XX Novel Notchless protein and nucleic acids encoding them useful for
 PT treating and preventing cancer and neurodegenerative diseases
 PS Claim 1; Page 44; 52pp; English.

XX The present sequence is that of Notchless, a novel protein of
 CC Drosophila. Notchless was identified in a screen for dominant
 CC modifiers of a Notch mutant phenotype in the Drosophila wing. The
 CC mutant dominantly suppressed the wing notching phenotype of
 CC notchoid mutations, and the Notchless protein was shown to bind to
 CC the cytoplasmic domain of Notch. Notchless modified Notch
 CC signaling activity in a variety of Notch-dependent signaling
 CC process in both Drosophila and Xenopus embryos. The Notchless
 CC protein has a novel highly conserved N-terminal domain followed by
 CC 9 WD40 repeats. Notchless, and nucleic acids encoding it, can
 CC be used in methods for the diagnosis and therapy of certain diseases,
 CC particularly cancer and neurodegenerative diseases (claimed). A
 CC Notchless mutant in a sensitised Notch genetic background is used
 CC in a claimed method for identifying Notch compounds capable of modifying
 CC the levels of expression or activity of a Notch protein.

SQ Sequence 479 AA;

Query Match 9.08; Score 229; DB 21; Length 479;
 Best Local Similarity 23.48; Pred. No. 1.3e-12;
 Matches 91; Conservative 50; Mismatches 128; Indels 120; Gaps 13;

QY 164 ALDDFYLNTLDWSSHNVLAVGLGNCVYLWNA-----CSSKYTKLDLGVDDCVCS 214
 DB 68 sledldiasvd--tenvidi-----yqgqavfkvrvtrctssmp-----ghaavvs 115
 QY 215 VGMAORGTHLAVGTNNGKVQIWDARCKIRSMDBGHLRYGALAMS--SSLSSGGGRKN 272
 DB 116 Infspdgahlasgsgdtlvtridntetphftctgkhqwlcvswapdqkrlasgcags 175
 QY 273 IYORIRTOEDFVSKLGSCKSEVCGIKW-----SYDNREILASGGDNKLFWNQHSSTOPY 327
 DB 176 lllwbelqgqgktrpilsghkhhincilawephrdpeckrlasasgddcriwklqgcl 235
 QY 328 LKYCEHTAAVKALAWSPHLHGLASGGGTADRCIRFWNTT----- 367
 DB 236 mlaaghtnavtavg--agliyle--sskdrtykmwraadgllcrlfsgnahvnnial 291
 QY 368 -----TNSHLS-----CMDTGSQVCNLY----- 385
 DB 292 sidvyltrpfrhvkdrskshlsisteelqesalkrygavcpd---eveslvsdndtl 348
 QY 386 --WSKVNELVSTHGYSON-----OIIVKRYPTMSKLAITLTGHT 422
 DB 349 ylvrmngnkcvemrghnqvndvkvyspdkllaasafkavrlwrasdgqymalfghv 408
 QY 423 YRVLYLAISPDOGTIVTGAGDETLRFWNV 451
 DB 409 gavytwawsadsrlivsgskdsclikwsv 437

Search completed: May 21, 2002, 11:46:21

Wed May 22 11:09:42 2002

us-09-701-572-2.rag

Page 16

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 11:44:12 ; Search time 20.97 Seconds
(without alignments)
2176.559 Million cell updates/sec

Title: US-09-701-572-2

Perfect score: 2555
Sequence: 1 MGCTGNRNPPTSTYRDNSP.....KSONTESEIGALSGRTTIR 475

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1839	72.0	444	2 T05122	hypothetical prote
2	1825	71.4	482	2 T09351	hypothetical prote
3	1086	42.9	702	2 T27730	hypothetical prote
4	1066	41.7	556	2 T37680	WD domain containi
5	985	38.6	421	2 T40614	G beta repeat prot
6	933.5	36.5	566	2 S64005	hypothetical prote
7	923	36.1	447	2 T05977	hypothetical prote
8	920.5	36.0	457	2 T05978	hypothetical prote
9	901.5	35.3	450	2 T14352	WD-repeat protein
10	900	35.2	499	2 B56021	probable cell divi
11	893.5	35.0	499	2 A56021	probable cell divi
12	876	34.3	440	2 T01768	hypothetical prote
13	777.5	30.4	488	2 T41719	CD20 protein homo
14	674	26.4	610	2 S64126	cell division cont
15	441.5	17.3	509	2 T41034	probable WD-domain
16	431	16.9	535	2 S62437	hypothetical prote
17	331.5	13.0	849	2 T49342	beta related protei
18	311	12.2	1526	2 AC2239	WD-40 repeat prote
19	310	12.1	336	2 T27762	hypothetical prote
20	278	10.9	676	2 AH2195	hypothetical prote
21	273.5	10.7	1747	2 AC1842	WD-40 repeat prote
22	273	10.7	1258	2 A12155	WD-repeat protein
23	272.5	10.7	786	2 AG3755	WD-40 repeat prote
24	272	10.6	1356	2 T18521	beta transductin-li
25	269.5	10.5	304	2 AG1837	WD-40 repeat prote
26	267	10.5	677	2 AE1861	serine/threonine k
27	256	10.0	1227	2 AE1810	WD-40 repeat prote
28	254.5	10.0	614	2 S58306	WD-40 repeat regul
29	250	9.8	586	2 T38992	WD-40 repeat regul

30	245	9.6	1189	2 A12493	WD-repeat protein
31	244.5	9.6	559	2 AB2202	hypothetical prote
32	244.5	9.6	1683	2 AF2071	WD-40 repeat prote
33	242.5	9.5	1711	2 AD1842	WD-40 repeat prote
34	238.5	9.3	265	2 AF1890	WD-repeat protein
35	237	9.3	515	2 S19487	hypothetical prote
36	234.5	9.2	1708	2 AE1866	WD-40 repeat prote
37	231	9.0	589	2 AC2400	WD-repeat protein
38	224.5	8.8	934	2 AG1889	trp-asp repeat con
39	221.5	8.7	502	2 T41148	hypothetical WD-re
40	221.5	8.7	777	2 T41075	hypothetical prote
41	221	8.6	473	2 T33805	probable u4/u6 sma
42	217.5	8.5	462	2 T50168	WD-repeat protein
43	215.5	8.4	1189	2 AH2154	beta transductin-li
44	213.5	8.4	1049	2 T42045	WD-repeat protein
45	211.5	8.3	598	2 AE2415	WD-repeat protein

ALIGNMENTS

RESULT 1

T05122

hypothetical protein F7H19.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 23-Apr-1999 #sequence_rev1stion 23-Apr-1999 #ext_change 26-May-2000

C:Accession: T05122

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.;

submitted to the Protein Sequence Database, July 1998

A:Reference number: 215399

A:Accession: T05122

A:Molecule type: DNA

A:Residues: 1-444 <BEV>

A:Cross-references: EMBL:AL031018

A:Experimental source: cultivar Columbia; BAC clone F7H19

A:Map position: 4

A:Introns: 140/3; 181/3; 216/3; 297/3; 362/3; 396/3

A:Note: F7H19.90

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match	72.0%	Score 1839;	DB 2;	Length 444;
Best Local Similarity	78.7%	Pred. No. 3.9e-135;		
Matches 350;	Conservative 39;	Mismatches 40;	Indels 16;	Gaps 7;
QY	22	PEPSESLRHYSRMINSNHYTSP-----SRTIYDFRIFPSRSASKALFDIN-TPT-EGR 74		
DB	4	PVLSIES--RINRLINANOSSPSSLSRSIYSDRFIPSRSGNFALFIDSPSPKDGK 61		
QY	75	DDSSAYTTLRTALFGDVAGPYTPKTDSPSMILPRNIFRKYKTRRSMHSLSPMD 134		
DB	62	EDGGSYATLTLRAAMFGE-----TPKRKDTGFS--SRNIFRKYKTRRSMHSLSPMD 115		
QY	135	DDFVGVNHS-PYKAPRKVPSPYKVIDAPALODDFYLNLDVMSHNVLAAGNCYTLW 193		
DB	116	DDDSGVHSGPVKAPRKVPSPYKVIDAPALODDFYLNLDVMSHNVLAAGNCYTLW 175		
QY	194	NACSSKVRKICLDLGVDCVCSVGAQGRTHLAVGTNNGKVOIWAACRCKTRSMGRHLR 253		
DB	176	NACSSKVRKICLDLGAEDSVCSVGAQGRTHLAVGTNNGKVOIWAACRCKTRSMGRHLR 235		
QY	254	VGALAWSSSLSSGGGRDNITORDIRQEDFVSKSGKRSVCGKMSYDNRRLASGND 313		
DB	236	VGALAWSSSVLSSGGGRDNITORDIRQEDFVSKSGKRSVCGKMSYDNRRLASGND 295		
QY	314	NKLFVWNGHSTQPVLYKCEHTAAVKAIAWSPHLLQGLASGGGTADRCIRFNNNTNHL 373		
DB	296	NKLFVWNGHSTQPVLYKCEHTAAVKAIAWSPHLLQGLASGGGTADRCIRFNNNTNHL 355		
QY	374	CMDTGSQVCNLAWSKNVDELSTHGYSONOIIIVARYPIPMKSLATLTGTTVAVLAIASPD 433		
DB	356	SDTGSQVCNLAWSKNVDELSTHGYSONOIIIVARYPIPMKSLATLTGTTVAVLAIASPD 415		

A:Cross-references: EMBL:AB005589; PIDN:BAA22152.1
 C:Genetics:
 A:Gene: str1L: SPAC144.13c
 A:Map position: 1
 C:Function:
 A:Description: may control differentiation and cell cycling by negatively regulating C1g
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 C:Keywords: cell cycle control

Query Match 41.7% Score 1066; DB 2; Length 556;
 Best Local Similarity 45.1%; Pred. No. 6.4e-75;
 Matches 223; Conservative 73; Mismatches 121; Indels 78; Gaps 10;

QY 42 TSPS-----RTYSDRFIPSRASAKFALDINTPTG-----RDDSSAYTTLRLTAL 89
 DB 61 SSPNNKRRRTMGDRFIPSRDAST-ELMTGFTKEGELTEVKKKQSVADRNTTLRLSEL 119
 QY 90 FG-----PDVAGPVTEKTDSPSMT-----LPN 112
 DB 120 FGSNDRTFNNSPIATPNTTIGVSPRTDSDIDIELTQRTPPSSSHSSILONTPTPS 179
 QY 113 RUIF-----RYKETROSMSLSPPMDDEFPVGNHSPVKAP---RK 151
 DB 180 KRIFFHLSRPNRKNSSYCKKAKQYQDNPRRTYSLP-----VRSITKDLISARLEGRE 233
 QY 152 VPRSPYKVLADALODDFYLLVDMSSHNVLAVGNCVYIMNACSSKVKIKDGLGVDDC 211
 DB 234 LPSIFRYLADAGLADDFYLLVDMSSHNVLAVGNCVYIMNACSSKVKIKDGLGVDDC 211
 QY 212 VCSVGAQGRTHLAVGTNNKGVQIWDARCKIRSMESGRHLRVGALAWSSSLSSGGGRD 271
 DB 294 VSLRWVQGTILAVGTNNKGVQIWDARCKIRSMESGRHLRVGALAWSSSLSSGGGRD 271
 QY 272 NIYORDIRTOEDFVSKLSGKSEVCGKMSYDNRELASGNDNKLFWNQHSTOPLVKYC 331
 DB 354 HILHRDVRAPRPFYFVLTFAHROEVCGLMNSNENLLASGNDNKLFWNQHSTOPLVKYC 331
 QY 332 EHFAAVKATAMSPHLHGLASGGTADRCIRFNTTNSHLSGMDTGSQVONLVMSKNV 391
 DB 414 NHTAAKATITWSPHOGIILASGGTADRCIRFNTTNSHLSGMDTGSQVONLVMSKNV 391
 QY 392 ELVSTHGSONQITVWRPYPTMSKATLTGHTYRVLYLAISPGQITVYAGDETLRFNNV 451
 DB 474 EPISTHGFMEVEVALMNPYSVSVGTGLKGTDRVLYLAMSPGENIVTGADETLRFNNV 451
 QY 452 FPPSPKSONTSEIGA 466
 DB 534 FDS-KSKHSASTWSS 547

RESULT 5

T40614
 G Beta repeat protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
 C:Accession: T40614
 R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, December 1998
 A:Reference number: Z21941
 A:Accession: T40614
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-421 <LTVN>
 A:Cross-references: EMBL:AL034563; PIDN:CAA22522.1; GSPDB:GN00067; SPDB:SPBC660.02
 A:Experimental source: strain 972h-; cosmid c660
 C:Genetics:
 A:Gene: SPDB:SPBC660.02
 A:Map position: 2
 A:Introns: 41/1
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 38.6% Score 985; DB 2; Length 421;
 Best Local Similarity 45.5%; Pred. No. 8.5e-69;
 Matches 196; Conservative 68; Mismatches 131; Indels 36; Gaps 7;

QY 51 DRFIPSRAS-----KRALF-----DINPTGRRDSSAYTTLRLTALFGPDVAG 96
 DB 3 DRFIPSRAS-----KRALF-----DINPTGRRDSSAYTTLRLTALFGPDVAG 96
 QY 97 P-----VTEPKTSPSMTLPNRIEFRTKTRQSMHSLSPMDDEFPVGNHSPVKAPR 150
 DB 61 SRATYGEDRKRIEKKMLDIPDRKSY-----SLSPISQSD-----MLRQPKPR 107
 QY 151 KVPSPYKVLADALODDFYLLVDMSSHNVLAVGNCVYIMNACSSKVKIKDGLGVDDC 210
 DB 108 AFPTPYKIIDAPYKIDDFYLLVDMSSHNVLAVGNCVYIMNACSSKVKIKDGLGVDDC 210
 QY 211 CVCSVGAQGRTHLAVGTNNKGVQIWDARCKIRSMESGRHLRVGALAWSSSLSSGGGRD 270
 DB 168 HVTSLMTGKGTOLAVGTDSGVITWIDESTKSVSLKGSERVAALAWMDNTLTSGGRD 227
 QY 271 NIYORDIRTOEDFVSKLSGKSEVCGKMSYDNRELASGNDNKLFWNQHSTOPLVKYC 330
 DB 228 EYLHHDLRAPGCCAEMKVEHDEICGLQMDRSJGLASGNDNKLFWNQHSTOPLVKYC 330
 QY 331 CERTAAVAVKATAMSPHLHGLASGGTADRCIRFNTTNSHLSGMDTGSQVONLVMSKNV 390
 DB 288 EHFAAVKATITWSPHOGIILASGGTADRCIRFNTTNSHLSGMDTGSQVONLVMSKNV 390
 QY 391 NELVSTHGSONQITVWRPYPTMSKATLTGHTYRVLYLAISPGQITVYAGDETLRFNNV 450
 DB 348 NELVSTHGSONQITVWRPYPTMSKATLTGHTYRVLYLAISPGQITVYAGDETLRFNNV 450
 QY 451 VF-PPSPKSONT 460
 DB 408 LENKRPKEEST 418

RESULT 6

S64005
 hypothetical protein YGI003c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein G3765
 C:Species: Saccharomyces cerevisiae
 C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 26-May-2000
 C:Accession: S64005
 R:Hehling, U.; Hofmann, B.; Delius, H.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64003
 A:Accession: S64005
 A:Molecule type: DNA
 A:Residues: 1-566 <HEB>
 A:Cross-references: EMBL:Z72525; NID:q1322450; PIDN:CAA96703.1; PID:e243314; PID:q132
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:CDH1
 A:Cross-references: SGD:S0002971; MIPS:YGI003c
 A:Map position: 7L
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 36.5% Score 933.5; DB 2; Length 566;
 Best Local Similarity 37.5%; Pred. No. 1.3e-64;
 Matches 209; Conservative 82; Mismatches 153; Indels 113; Gaps 13;

QY 6 NRPNPPTSTVADNSPPPEP--SPESLRHYSRMINSNH---YTSPEP-----TIYSDRPI 54
 DB 4 NLNF-----FMNNTPPSSPLKGSSEKRSKRPYSSSSASLSSPSRRSRPSTVYGDRTY 58
 QY 55 PSR-----SASKRALDINTPTGGRD-----DSSAYTTLRLTALFG----- 91
 DB 59 PSRTDIDFNSIVSISMAVPAALPSPSTEDOVEROKERQAHETRYNTLLKNELFGEMLSKD 118
 QY 92 -----PDVAG-----PYTPE 101

```

Db 119 TVGSESSIDRIKNTSPSTGCVNHAENTTRHGVELEFSTPPEAAGLEEFSPHSTPYPR 178
QY 102 K---TDSPSMTLPNNIR-----YKTE-----TROSMHSLSPMDDEFPVPGV 141
Db 179 RLFTSQDEITRPSNSNRGASLTYOQRGRRLSASLQSOFEFSSPVRPDS--KQL 236
QY 142 NHPVKARVPRSPYKVLADAPALQDDFYLNLVDMSSHNLAVGAGNCVYLMNACSSKYT 201
Db 237 LLSPEKQFQIAKYRYVRLADPVLADDFYSLIDMSSDVLAVLALGSLFILDNTGDDV 296
QY 202 KLCDDGVDDCYGVMQORGHILAVGNNKVOIWDARCKKIRSMGHLRGLVAMSS 261
Db 297 HLCQ--TENETYSLSMIGAGSHLAVGQANGLVEIYDWKRCRITLGLHIDRVACLSMNN 354
QY 262 SLSSGGDKNIYORDIRTOEDFVSKLSGHSVEYGLKWSYDNEELASGGNDKTLFVWQ 321
Db 355 HVLTSGRDHRILHRDVRMPDPFETIESHQEYCGIKMVAADKGLASGGNDVNVHYEG 414
QY 322 HSTQPVLYKCEHTAAVAIAVMSPHLGLASGGGTADRCIRFNTTNSHLSGMDTGSQY 381
Db 415 TSKSPILTFDEHKAAYAMAMSPHKGVLATGGGTADRLKIMNVMSIKMSIDSGSQI 474
QY 382 CNLWMSKNVIELSTHGYSONOIIWYRPMKSLATLGHYRVLYLAISPDGQITVGA 441
Db 475 CNMWSKNTMELVTSRYSKYNLTLMDCNMDPLAILKGHSFVRLHLLTSLNDGTIVVSGA 534
QY 442 GDETLRFWNVPSPKSQ 458
Db 535 GDETLRFWKLDFDKPAK 551

```

RESULT 7

hypothetical protein F17M5.20 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 26-May-2000
 C:Accession: T05977
 R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hobeisel, J.; Mewes, H.W.; Mayer, K.F.X
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15263
 A:Accession: T05977
 A:Molecule type: DNA
 A:Residues: 1-447 <BEV>
 A:Cross-references: EMBL:AL035678; GSPDB:GN00062; ATSP:F17M5.20
 A:Experimental source: cultivar Columbia; BAC clone F17M5
 C:Genetics:
 A:Gene: ATSP:F17M5.20
 A:Map position: 4
 A:Introns: 23/3; 116/3; 206/3; 400/3
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 36.1%; Score 923; DB 2; Length 447;
 Best Local Similarity 41.1%; Pred. No. 6; le-64;
 Matches 196; Conservative 87; Mismatches 124; Indels 70; Gaps 11;

```

QY 1 MDGTGNRPPTSTVRDMSPPPEPSRLRHSRMINSNHTSPRTIYSDREIFPSAS 60
Db 1 MDAGINRCP-----LQEHFLPRKNSKNTL-----DRLPNSAM 34
QY 61 KF--ALPDINPTGRDSSAYTTLTALFQPDVAGVTP--EKTDSPSMTLPNNIF 116
Db 35 NEDVAHFALTEERKQDQASAT-----VSSPSKEAYRKQLAETMNLNHRIL 80
QY 117 RYKTETRSMHSLSPFMDDEFPVGNHS-----PYKAPRKVPSPRYKVLADPALQDD 168
Db 81 AFRNK-----POAPVELLP--NHSASLHQOPKSKPRRYIPQTSERTLAPDIDVD 130
QY 169 FYLLLVDMSSNVLAVGLGNCVYLMNACSSKYTKLCDLGYD--DCVCSVGAQRTHLAVG 227
Db 131 FYLLLVDMSSANVLAIDLHYVLMDSATGSTSELVITDEKQYVSIINAPDGRHVAVG 190
QY 228 TNNCKVOIWDARCKKIRSMG--GHLRLVGLAMASSLSLSSGGDKNIYORDIRTOEDFVS 286

```

```

Db 191 LNNSEYQIMDSASNRQLRTLKGGHQSRLAMNNHILTTGGMDGLINNDVIRIRPIYE 250
QY 287 KLSGHSYVGLKWSYDNRRLASGGNDKLFVWQ-----HSTQPVLYKCEHTAAVAI 340
Db 251 TYRGHTQEVCGLWMSGSGQOLASGGNDVYHIDRSVASSNSTQWLHRLHEHTSAVKAL 310
QY 341 AMSPHLHGLASGGGTADRCIRFNTTNSHLSGMDTGSQVNCNLVMSKNVNEVSHGYS 400
Db 311 AMCPQANILATGGGGDRTIKFNNHTGACLSNVDTSQVCSLMSKNERELSSHGFT 370
QY 401 QNQLIYWRPYMSKSLATLGHYRVLYLAISPDGQITVGADETLRFWNVPSPKS 457
Db 371 QNQLTLKPYSMVKMELTGHYSRVLYMAQSPDGTVAAGDETLRFWNVGVPE 427

```

RESULT 8

hypothetical protein F17M5.30 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 26-May-2000
 C:Accession: T05978
 R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hobeisel, J.; Mewes, H.W.; Mayer, K.
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15263
 A:Accession: T05978
 A:Molecule type: DNA
 A:Residues: 1-457 <BEV>
 A:Cross-references: EMBL:AL035678; GSPDB:GN00062; ATSP:F17M5.30
 A:Experimental source: cultivar Columbia; BAC clone F17M5
 C:Genetics:
 A:Gene: ATSP:F17M5.30
 A:Map position: 4
 A:Introns: 33/3; 126/3; 216/3; 410/3
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 36.0%; Score 920.5; DB 2; Length 457;
 Best Local Similarity 41.8%; Pred. No. 9; le-64;
 Matches 191; Conservative 84; Mismatches 117; Indels 65; Gaps 10;

```

QY 21 PPEPSPELRHVSRMINSNHTSPRTIYSDREIFPSASKF--ALPDINTPEGRDSS 78
Db 26 PRKPSKENTL-----DRLPNSAMNPFYAHFALTEGRKQDQTA 64
QY 79 SAYTTLRTALFQPDVAGVTP--EKTDSPSMTLPNNIRIYRTETROSMSLSPPMDD 136
Db 65 A-----VSSPSKEAYRKQLAETMNLNHRILAFRNK-----POAPVE 101
QY 137 FVPGVNH-----PYKAPRKVPSPRYKVLADPALQDDFYLNLVDMSSHNLAVGLGN 188
Db 102 LKPS--NHSASLHQOPKSKPRRYIPQTSERTLADPVIDDFYLLNDWGSANVLAIALDH 160
QY 189 CYLLMNCSSKYTKLCDLGYD--DCVCSVGAQRTHLAVGNNKVOIWDARCKKIRSM 247
Db 161 TVYLMDSSTSTSELVITDEKQYVSIINAPDGRHVAAGLNSEYQIMDSASNRQLRTL 220
QY 248 E--GHLRLVGLAMASSLSLSSGGDKNIYORDIRTOEDFVSKLSGHSVEYGLKWSYDNE 306
Db 221 KGHQSRVGSLLAMNNHILTTGGMDGLINNDVIRIRPIYETRGHTQEVCGLWMSGSGQO 280
QY 307 LASGGNDKLFVWQ-----HSTQPVLYKCEHTAAVAIAVMSPHLGLASGGGTADRC 360
Db 281 LASGGNDVYHIDRSVASSNSTQWLHRLHEHTSAVKALAMCPQANILATGGGGGRT 340
QY 361 IRFNNNTTNSHLSGMDTGSQVNCNLVMSKNVNEVSHGYSOQIYWRPYMSKSLATLFG 420
Db 341 IKFNNHTGACLSNVDTSQVCSLMSKNERELSSHGTOQLTKMYPVKMAELTG 400
QY 421 HTRVLYLAISPDGQITVGADETLRFWNVPSPKS 457
Db 401 HTRVLYMAQSPDGTVAAGDETLRFWNVGVPE 437

```

C;Superfamily: unassigned WD repeat proteins; WD repeat homology
F;305-338/Domain: WD repeat homology <WD1>

1

1

1

```

Db      22 NAPPARWOKKAKKAAGPAPSPMRAANSHSAGRTPGRTPGKSSSKVOTTPSKP-GGDRY 80
QY      55 PRRSASKFALPDINTPTTEGRDSSATTLTALFGLPDVAGPYTEKTDSPSMTLPNNR 114
Db      81 PRRSAAQMEVASFILSKENQSENQ-----PTKKHOKAMALNIN 121
QY      115 -----IFRYKTTETROS-----MHSLSPEMDDEFPVGNHSPYKAPKRYPSPKYVDA 162
Db      122 GFDEVEAKILRLSGKPPQNAPEGYONRLKVLYSOKATPG---SSKTCRYTFLPRLIDDA 178
QY      163 PALODDEYLVNPMSSNNVLAVALGNCVYIMANACSKVTKLCDL-GVDDCVCVGMAGQ 221
Db      179 PEIINDVYLVNPMSSNNVLAVALDMSVYLSMSASGDIQLQMEQPEYISSVAMIREG 238
QY      222 THLAVGTNNKQVQIMDAARCKKIRSMEGHRLRYGALAMSSLLSSGGDKNIYODITQ 281
Db      239 NYLAVGSSSAVQVLMVQOQKRLNMTSHSAVSGLSMNSYLLSSGSHHHHVDVVA 298
QY      282 EDFVSKLISGHSVGLKMSYDNRELASGDNKLFVW---NQHSTOPVLKYEHTAAV 337
Db      299 EHHVATLSGHSOEVGCLMAAPDGRHLASGNDLVNVPSPAPGEGWVPLQTFQHOQAV 358
QY      338 KAIAMSPHLGLASGGGTADRCIRFMMNTTNSHLSGMDTGOYCNLYMSKNNVELYTH 397
Db      359 KAVAMCPQSNVLAATGGGSTRHIRMVCGACLSANDASOVCSTLMSPHYELISGH 418
QY      398 GVSQNOIYVWRPTMSKLTATLGHTRYLYLAISPQGTITVGADETLRFMMVFPSPKS 457
Db      419 GFAQNOIVYWKPTMAKVAELKSHTSVSLTMSPDGATVSAADETLRLMRCFELDPA 478
QY      458 QMTESE 463
Db      479 RRRERE 484

```

RESULT 12

hypotheoretical protein A_IG002P16.8 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 26-May-2000
 C:Accession: T01768
 R:Miller, N.; Beck, C.; Kramer, J.
 Submitted to the EMBL Data Library, June 1997
 A:Description: The sequence of A. thaliana IG002P16.
 A:Reference number: Z14421
 A:Accession: T01768
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-440 <MIL>
 A:Cross-references: EMBL:AF007270; NID:g2191157; PID:g2191163; GSPDB:GN00063; ATSP:A_IG002P16.8
 A:Gene: ATSP:A_IG002P16.8
 A:Map position: 5
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 34.3%; Score 876; DB 2; Length 440;
 Best Local Similarity 40.6%; Pred. No. 2.7e-60;
 Matches 189; Conservative 74; Mismatches 129; Indels 74; Gaps 10;

```

QY      12 TSTVDNSPPPEPESLRLHVSRLMNSNHTSPSRTIYSRPTPSRSASFALPDINTPT 71
Db      4 TCTVPEHFLPKRLSKONL-----DRETPNRSKAFED-FANVYALT 41
QY      72 EG-----RDDSSAYTTLTALFGLPDVAGPYTEKTDSPSMTLPNNR-----IFRY 118
Db      42 OGSKRNDEVTASAKKAYTQALAVM-----NONRRIILAFNN 79
QY      119 KTETRQSK-HSLSPFMDDEFPVGNHSPYKAPKRYPSKYVDAALODDEYLVNPM 177
Db      80 KPKSLSTNHSDDPHQ-----NRPYKPRRYTIPQNSERVLADAGLDDSLNLLDMG 131

```

```

QY      178 SHNVLAAGLNCVYLMANACSKVTKLCDLVD-DCVCSVGMAGRTHLAVGTNNKQVQ 236
Db      132 SANVLAALAGLTVYLMADSSGSGSTSELTIDEDKGPVTSIMWODGLDVLAVGDNSEVQ 191
QY      237 DAARCKKIRSMW-GHRLRYGALAMSSLLSSGGDKNIYODITQEDFVSKLSGHKSEV 295
Db      192 DCVSNQVRLRLRGHESRVSGLAMDNHLLTQGMQDKIYNNQVRLRSISIVETYLGHTEBY 251
QY      296 GGLKSTNDNELASGDNKLFVW-----NQHSTOPVLKYEHTAAVAKIAMSPLHLGL 349
Db      252 GGLKWSSEGNKQASGNDNVVHIMDRSLASSKOTRQMLHREFEHTAAVALAMCPQASL 311
QY      350 LASGGGTADRCIRFMMNTTNSHLSGMDTGOYCNLYMSKNNVELYTHGYSQNOIYWR 409
Db      312 LATGGVGQDGKIKFMMTHTGACLSVETGSOVCSTLMSQSEHELLSHGFTQNDLWKY 371
QY      410 PTMSKLTATLGHTRYLYLAISPQGTITVGADETLRFMMVFPSP 455
Db      372 PMSKMAELNGHSTRVLFMAQSPNGCTVASAAGDENLRLMVFGEPR 417

```

RESULT 13

CDC20 protein homolog - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
 C:Accession: T41719
 R:Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
 Submitted to the EMBL Data Library, October 1999
 A:Reference number: Z22012
 A:Accession: T41719
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-488 <RIE>
 A:Cross-references: EMBL:AL121770; PID:CA557442.1; GSPDB:GN00066; SPDB:SPAC821.08c
 A:Experimental source: strain 972h; cosmid c821
 A:Gene: SPDB:SPAC821.08c
 A:Map position: 1
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 30.4%; Score 777.5; DB 2; Length 488;
 Best Local Similarity 37.4%; Pred. No. 1.4e-52;
 Matches 181; Conservative 73; Mismatches 169; Indels 61; Gaps 12;

```

QY      2 DGTGNMNPPTSTVRDNSPPPE-----PSESRLHVSRLMNSNHTSPSRTIYSDF 53
Db      42 NGRSSKRCSFKSSFTIRNSPKIDVNTDMSIPLCGSPRKRK-----PASR---SDRF 90
QY      54 IPSR--SASKFALPDINTPTTEGRDSSAYTTLTALFGLPDVAGPYTEKTDSPSMTLP 111
Db      91 IPSRPTANAF-----VNSIS---SDVPDYSESAEAC-GFDLNTVLAFLAFLAPKPKP 142
QY      112 NNIRFRYKTEFQSMHSLSPEMDDEFPVGNHSPYKAP--RKYPRSPKYVDAALODDE 169
Db      143 ---VDLRTQHNRRPQ-----PVYPAKRFRFTPTPERVLDAPIIIDY 181
QY      170 YLVNLMSSHHVLAVALGNCVYLMANACSKVTKLCDLGVDDCVCVGMAGRTHLAVGTN 229
Db      182 YLVNLMSSNNVLAVALERNTYVNNADSGVSAALAEDESTYVAASVKSWSHDSGLSVGIG 241
QY      230 NGKQVQIMDAARCKKIRSMEGHRLRYGALAMSSLLSSGGDKNIYODITQEDFVSKLS 289
Db      242 NGVDIYDYESQTKLRTMAGHQARVGLCLSNRRHVLVSSGSRGAIHHHVDRIANHOIGTLQ 301
QY      290 GHRSEVGLKMSYDNRELASGDNKLFVWQNSHSTOPVLKYEHTAAVAKIAMSPLHLGL 349
Db      302 GHSEVCGLAMRSDGLOLASGNDNVVQIMDASSIFPKFTKTNHNAAYAVAMCQWQNL 361
QY      350 LASGGGTADRCIRFMMNTTNSHLSGMDTGOYCNLYMSKNNVELYTHGYSQNOIYWR 409
Db      362 LATGGGTMDKQIHFWMAATGARVNTVDASQVTLMSPHSKELMSTHGFPDNNLSIWY 421

```

276 AIALDT
Db

Db 2/6 ALAEDINUL.....

Wed May 22 11:09:48 2002

Search completed: May 21, 2002, 11:46:48
Job time: 156 sec

us-09-701-572-2.rpt


```

Db 4 NLRP-----FMNTPSSSPLKGSSEKSKRVSKRPISSSSSASLSSPSSRRSPSTVYGDYI 58
QY 55 PSR-----SASKFALFDINPTGRD-----DSSAYTLLRLTALFG-----91
Db 59 PSRTIDPNSIVISISSMASVAPALNPSTEDOVYOKROKHEFYNTLLNLELFGELSLD 118
QY 92 -----PVAG-----PVPE 101
Db 119 TVGSSSIDRIKTRPSTRGVNAENTTRHGYELERVSTPPEDAGLEEFSPHSTPYPR 178
QY 102 K---TDSPTMTLPNRNFR-----YKTE-----TRQSHSLSPFMDDEFPYGV 141
Db 179 RLFTSODDELTTPSSNSVRSASLITYOKRKRLSASLLOSOFDMSFPRPDS--KOL 236
QY 142 NHPVKAPRKVRPSRYKVLDAPALDDFYLVMSHNVAVGNCVYIMNACSSKVT 201
Db 237 LLSFGOFROIAPVPRVADAPSLADDFEYSLIDMSSTVLAVALGSIETLDNNTGDVY 296
QY 202 KLCDLGVDDCVCSVGAQGTHTAVGTNNKVQITDAARCKIRISMEGHRRLVGLAMSS 261
Db 297 HLCD--TENEYSLSWIGAGSHLVAGQANGLVEIYDVKRCKITLTSGLHIDVACLSMNN 354
QY 262 SLLSGGRDKNIYORDITQEDFVSKLSGKSEVCGKMSYDNRBLASGDNKKLFVMO 321
Db 355 HVLTSGSNDHRLHLDVMPDPFEETTESHTQEVCGKLVADNKNVAGNDVAVHYEG 414
QY 322 HSTOVLYKCEHTAAVNAIAPLHGLLASGGTADRCIRFVNTTNSHLSGMDTGSOV 381
Db 415 TSKSPILTFDEHKAAYKAMAMSPHAKRVLATGGGTADRRKLIKIMNVTISIKMSDIDSGOI 474
QY 382 CNLWMSKVNELVTHGYSOQITVRYPTMSKRLATLGTFRVYLYLISPGQOTYVGA 441
Db 475 CNMWSKNTMELVTHGYSKYNLTLDWCNSMDPIALIKGHSFRLVHLTSLSDGTTVSGA 534
QY 442 GDETLRWNVFPPSKQ 458
Db 535 GDETLRWKLEDKPKAK 551

```

RESULT 2
CC20 YEAST
ID CC20 YEAST STANDARD: PRT: 610 AA.
AC P26309;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division control protein 20.
GN CDC20 OR YGL116W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
NCBI_Taxid=4932;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-A364A;
RX MEDLINE=92017840; PubMed=1922065;
RA Sechi N., Montegudo M.C., Koshland D., Hogan E., Burke D.J.;
RT "The CDC20 gene product of Saccharomyces cerevisiae, a beta-transducin
RT homolog, is required for a subset of microtubule-dependent cellular
RT processes";
RL Mol. Cell. Biol. 11:5592-5602(1991).
RN
RP SEQUENCE FROM N.A.
RA Lauguin G.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RP
RA
RA Doi A., Doi K.;
CC
CC -1- FUNCTION. REQUIRED FOR TWO MICROTUBULE-DEPENDENT PROCESSES.
CC NUCLEAR MOVEMENTS PRIOR TO ANAPHASE AND CHROMOSOME SEPARATION.
CC CDC20 MAY MODULATE MICROTUBULE STRUCTURE EITHER BY PROMOTING
CC MICROTUBULE DISASSEMBLY OR BY ALTERING THE SURFACE OF THE

```

CC  
CC MICROTUBULES.  
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
CC -1- CAUTION. REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 502  
CC ONWARD AND IS SHORTER (519 AA) DUE TO A FRAMESHIFT.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch)  
CC  
CC EMBL: X59428; CAA42058.1;  
CC EMBL: 272638; CAA96824.1;  
CC EMBL: D16506; BAA03957.1;  
CC PIR: S18840; S18840.  
CC SGD: S0003084; CDC20.  
CC InterPro: IPR000002; Flazy.  
CC Pfam: PF00400; WD40; 5.  
CC ProDom: PD004563; Flazy; 1.  
CC SMART: SM00320; WD40; 5.  
CC PROSITE: PS00678; WD_REPEATS_1; 4.  
CC PROSITE: PS50082; WD_REPEATS_2; 2.  
CC Cell division: Microtubules; Repeat; WD repeat.  
CC  
CC REPEAT 257 296  
CC REPEAT 299 338  
CC REPEAT 342 379  
CC REPEAT 383 422  
CC REPEAT 425 467  
CC REPEAT 469 519  
CC REPEAT 523 562  
CC CONFLICT 318 319  
CC CONFLICT 502 519  
CC  
CC IC -> MA (IN REF. 3).  
CC PENALTYVETKFAE -> QMOSLIIMONSKLOK  
CC (IN REF. 1).  
CC MISSING (IN REF. 1).  
CC  
CC CONFLICT 520 610  
CC SEQUENCE 610 AA; 67359 MW; CCE7CD1A9C15ACF CnC64;  


```

Query Match 26.4%; Score 674; DB 1; Length 610;
Best Local Similarity 31.7%; Pred. No. 1.7e-45;
Matches 151; Conservative 87; Mismatches 164; Indels 74; Gaps 13;

```

QY 6 NRPPTSVRNSPPEPSPESLRHVSRLMNSNHTYSPRTIYSDRFPSASAKFALF 65
Db 154 SQNKVDPELTLEALPVPNAPSIS-----HLRAQTKIYF-----KQVMAEAGL 196
QY 66 DINTPEGRDSSSAVYTLRTALFPGDVAGPYTPKTDSPSTLPNRIFFRYKTET 122
Db 197 DMNK-----RILOYME-----PKCSLSRQSYIMKKRFRHYSQOEO 234
QY 123 ROSHMSLSPEFMDDEFPYGVNHSVPKAPRKVRPSRYVLDAPALDDFYLVMSHNV 182
Db 235 K-----IPDL-----IKL-RKINTPREIIDAPGFODFTLNLSSKKNVL 275
QY 183 AVGAGNCVYIMNACSSKVTYKLCGLDGVDCVSGVNAQGTHTAVGTNNKVQINDAARCK 242
Db 276 AIALDTALVLYMNAITGDVSLTDF-ENTTICSVTSDDCDHISIGKEGNTETIDVETMS 334
QY 243 KIRSM-EGHRLRYGALAWSSLSLSSGGRDKNIYORDITQEDFVSKLSGKSEVCG 301
Db 335 LIRTRSGGLGVIRGISLWLDLITATGSRSGEIQINDVIRKOHIVSTAHEITGVCGLSIK 394
QY 302 YDNRELASGDNKKLFVNNQSTQPVLYKCEHTAAVNAIAPLHGLLASGGTADRCI 361
Db 395 SGOLOLASGNDNTYMWIDRTSLPQPSKTHPAVAALSWCPYSPNITLASGGGQTDKHI 454
QY 362 RFWNTTNSHLSGMDTGSOVNCLWMSK-----VNELVSTHGYSONOITVRYPTM 412
Db 455 HFWNSITGARVSINTGSVSLHWGQSHSTNGCMNKNKEIVATGNGPENALSVNYETK 514

```

OY 413 SKLATLT-GHTYRLVLAISPDCQITVAGDETLAEFNNPPSPKQNTSEIGAL 467
 DB 515 FKVAEVAHARFICCSQSPDGTTLATVGGDENLKFYIF-DPROTGSRBDIGM 569

RESULT 3
 Y98_SCHPO STANDARD: PRT: 535 AA.
 ID Y98_SCHPO
 AC 009786;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 60.5 kDa Trp-Asp repeats containing protein C1366.08 in
 chromosome 1.
 GN SPAC1366.08.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 CC NCBI_TaxID=4896;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=972;
 RA Odell C., Bowman S., Bartell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z54308; CAA91101.1;
 CC InterPro: IPR000002; F1zyz.
 CC InterPro: IPR001680; WD40.
 CC Pfam: PF00400; WD40; 5.
 CC ProDom: PD004563; F1zyz; 1.
 CC ProDom: PD00320; WD40; 3.
 CC SMART: SM00320; WD_REPEATS_1; FALSE_NEG.
 CC PROSITE: PS00678; WD_REPEATS_1; FALSE_NEG.
 CC PROSITE: PS00678; WD_REPEATS_2; 2.
 CC PROSITE: PS00678; WD_REPEATS_REGION; 1.
 CC PROSITE: PS00678; WD_REPEATS_REGION; 1.
 CC Hypothetical protein; Repeat; WD repeat.
 CC KW REPEAT
 CC FT REPEAT 189 226 WD 1.
 CC FT REPEAT 228 267 WD 2.
 CC FT REPEAT 269 314 WD 3.
 CC FT REPEAT 320 359 WD 4.
 CC FT REPEAT 362 404 WD 5.
 CC FT REPEAT 462 505 WD 6.
 CC FT REPEAT 535 AA; 60533 MW; F438AE5981C65B42 CRC64;
 SO SEQUENCE

Query Match 16.9%; Score 431; DB 1; Length 535;
 Best Local Similarity 27.8%; Pred. No. 2e-26; Indels 112; Gaps 20;
 Matches 142; Conservative 80; Mismatches 176;

OY 12 TSTV-----RDNSPP-PEPSPESLRHYSRMINSNHTYSPRTIYS-----DRF 53
 DB 30 SSTVYGMNERNTPLTIDPSKNETANVC-PIKSLFQSYKSGYSRSQRLTSRNGIDR 88
 OY 54 IPSASAKFALFDINTPEGRDS-----SSAYTTLRLALFGPDVAGPYTPE 101
 DB 89 IPMTS-----NKDTISGRHSLSRLNKNKNASETYOOLEYAL-----EVE 132
 OY 102 KTDSPMTLPKRNIFKKTETROSMSLSPFM---DDDFVPGVNSPVKAPKYPKSPYK 158
 DB 133 RDD-----NVTYAKLOKSDMOKCPTVASEKDKGNLKNKNSPENTL--LDFR 181
 OY 159 VLDAPALODDFYLNLDVMSHNVLAVGLNGCYVLNNAKSKYTKLCCDGLVDCCVCSGMA 218
 DB 182 IIDAPELRDFYTLSSLSWSPKGLAIGLAENIYLMK-ELGPRVLEESITYD-VSSVAYS 239

OY 219 ORGTHLANGTNGKVOIN-DAROCKIRSMGHRRLRYGALAMSSSL-----LSGGRDNI 273
 DB 240 YNDILAVGRVYGLTLOFQODNERPRI--HHPGDIGVLAHMKRVLETNLLGKGNMI 297
 OY 274 YORDIRTOEDFYSK-----LSGKSEYCGLKMSYNDRELASGDNKLFVNNHSTOP 326
 DB 298 FYVDIIMSES-TSKAVLVATITNAHDEQVCLTNHNSGOFASGDNRCLEFGSDLRQ 356
 OY 327 VLKXCEHTAAVKAHIAHSLHGLASGGTADRCIRFNTTNSHLSGMDTGSQVCLVW 386
 DB 357 PLYWQONAAVKAHSLFPCWQORSILATGASGSHKHIFVNCNKGKIDELVCGAQTITSH 416
 OY 387 SKANNELVSTHGS-----ONQIIVWRYPMTSKLATLTGHTYRLVLAISP- 433
 DB 417 SPKREFSVTFGVSLFVQHRFAVYSMPOLCVSV-----LPEVPDIRCVHSVLT 467
 OY 434 -----GGTYTGAGDETLAEFNN 451
 DB 468 SOLNETTGRCMSTDSITIASNETIKFEDL 497

RESULT 4
 PKWA_THECU STANDARD: PRT: 742 AA.
 ID PKWA_THECU
 AC P49695;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Putative serine/threonine-protein kinase PKWA (EC 2.7.1.-).
 GN PKWA OR PKM1.
 OS Thermomonospora curvata.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangiales; Thermomonosporaceae;
 OC Thermomonospora.
 CC NCBI_TaxID=2020;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CCM 3352;
 RC MEDLINE=96200125; PubMed=8631732;
 RA Janda L., Tichy P., Spizek J., Petrlick M.;
 RT "A deduced Thermomonospora curvata protein containing
 RT serine/threonine protein kinase and WD-repeat domains";
 RL J. Bacteriol. 178:1487-1499(1996).
 CC -1 FUNCTION: MAY PLAY A REGULATORY ROLE DURING THE COMPLEX GROWTH
 CC CYCLE AND IN SECONDARY METABOLITE PRODUCTION.
 CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1 SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF115313; AAB05822.1;
 CC HSPSP; P00523; 2PTR.
 CC InterPro: IPR000719; Euk_pkinase.
 CC InterPro: IPR002290; Ser_thr_pkinase.
 CC Pfam: PF00069; pkinase; 1.
 CC Pfam: PF00400; WD40; 7.
 CC PRINTS: PR00320; GPOTEINBRPT.
 CC SMART: SM00320; WD40; 7.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00108; PROTEIN_KINASE_DOM; 1.
 CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 CC PROSITE: PS00678; WD_REPEATS_1; 5.
 CC PROSITE: PS00678; WD_REPEATS_2; 7.
 CC PROSITE: PS00678; WD_REPEATS_REGION; 1.
 CC Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat;

KW WD repeat. 16 266 PROTEIN KINASE.
 FT DOMAIN 22 30 ATP (BY SIMILARITY).
 FT NP_BIND 22 30 ATP (BY SIMILARITY).
 FT BINDING 44 44
 FT ACT_SITE 138 138
 FT DOMAIN 301 394
 FT REPEAT 455 496
 FT REPEAT 497 538
 FT REPEAT 539 580
 FT REPEAT 581 621
 FT REPEAT 622 663
 FT REPEAT 664 705
 FT REPEAT 706 742
 SO SEQUENCE 742 AA; 78950 MW; AC173460DB4383D CRC64;

Query Match 11.4%; Score 290; DB 1; Length 742;
 Best Local Similarity 25.7%; Pred. No. 3.9e-13;
 Matches 123; Conservative 71; Mismatches 199; Indels 86; Gaps 20;

QY 12 TSTVRDNDSPPEPSPELRLHVSRLMINS---NHYTSPSRITYSDRFPSRSASKFALFDI 67
 DB 266 THRTGDMDPP-----HLSDLIGSMPLGATTSNPSLALEPPPSHGPPR----- 311
 QY 68 NPTTEGRDSSAYTTLRTALFGPDVAGVPTPEKTSPEMTLPNNRIERYKTETROSMH 127
 DB 312 --PSEPLDPP-----GDDADP-----SAKPSRTLPPEPEPELEKPIQVTH 352
 QY 128 S---LSPPMDDFVGVNHSVPKAPKRVPSPKVLADALDDFVLNLVDMSS-----H 179
 DB 353 EPERAPTPRRPREPARGATKPKNPR--PAAPQPPWSPRPVQPPRMKQILTKKPAVAGILT 410
 QY 180 NVLAVGNCVYLMN-----ACSSKVTIKCD-----LGVDDCVCSVGMQR 220
 DB 411 AVATAGLVSVFLVWQMTLPETPLRPDSSTAPSESADPHNLNPRLLTDRRAVAVAFSPG 470
 QY 221 GTHLAVGTNNCKVQIWDAAKCKIRSMGHRIRVGLALMS--SSLSSGGDKNIYORDI 278
 DB 471 GSLAGSGDGKLIHWDVAGSDELTHTLBSGHTDWRVAVAFSPDGLALASGDDATVRLMDV 530
 QY 279 RTGEDFVSKLSGKSEVCLAKWSYNRELASGNDNKFVWN-----OHSTQPVLAKEH 333
 DB 531 AAEEF-RAVPEGHTHYVDIAFSPDGSVAVASGRGTARLNNVATGEHA---VLK--GH 584
 QY 334 TAAVKAIAVSPHLGLLASSGGGTADRCIRFNNNTTNSHLSMDTGSQ--VCNLVMSKNVNE 392
 DB 585 TYTYVAVAFSPD--GSMVAS--GSRDGTIRLMDVATGKERDYLQAPENNVSLAFSPDGS 641
 QY 393 LSTHGYSONOIIWRYRYTMSKLAITLGHTRYVLYLAISPDGQIYVAGDETLRFNMY 451
 DB 642 LV--HG-SDSTVYHMDVAVASGEALHTFEHGDWVRAVAFSPDGLALASGSDDTIRLMDV 697

RESULT 5
 YSOO_ANASP STANDARD; PRT; 1258 AA.
 AC OBYTC2;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical WD-repeat protein alr2800.
 GN alr2800.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 NX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Matsumoto A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;

"Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120.",
 DNA Res. 8:203-213(2001).
 --- SIMILARITY: CONTAINS 15 WD REPEATS (TRP-ASP DOMAINS).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AP003590; BAB74499.1; -
 DR PROSITE; PS00678; WD_REPEATS.1; 9.
 DR PROSITE; PS50082; WD_REPEATS.2; 14.
 DR PROSITE; PS50294; WD_REPEATS_REGION.1.
 KW Hypothetical protein: Repeat; WD repeat; Complete proteome.
 FT REPEAT 55 93
 FT REPEAT 640 679
 FT REPEAT 682 721
 FT REPEAT 724 763
 FT REPEAT 766 807
 FT REPEAT 809 849
 FT REPEAT 850 889
 FT REPEAT 892 931
 FT REPEAT 934 975
 FT REPEAT 976 1017
 FT REPEAT 1019 1059
 FT REPEAT 1060 1101
 FT REPEAT 1103 1143
 FT REPEAT 1144 1183
 FT REPEAT 1186 1227
 SQ SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;

Query Match 10.7%; Score 273; DB 1; Length 1258;
 Best Local Similarity 29.9%; Pred. No. 1.8e-13;
 Matches 81; Conservative 54; Mismatches 126; Indels 10; Gaps 8;

QY 184 VGLNCYVLYNNACSKVTKLCDLGVDDCVCSVGMQRGTHLAVGNCKVQIWDAAKCK 243
 DB 912 VSLDOAVRLNCCRTGOCCLKAK--YGNIDMALPYAFSPDQILASGNSKRYVLTMDMQTKY 970
 QY 244 IRSMEGHRIRVGLALMS--SSLSSGGDKNIYORDITFODFVSKLSGKSEVCGCLKWS 301
 DB 971 ISSLEGHTEPIYGIASFPSDQTLASASTDSSVRLMNISTGQCFILLE-HTDWYAVAFH 1029
 QY 302 YDNRELASGNDNKLFWNQHSTQVLEKCEHTAAVKAIAVSPHLGLLASSGGGTADRCI 361
 DB 1030 PCKIATGSACTVAKLWNISITGOCCLKTLSHSDKILGMASPD--GQLLAS--ASADQSV 1086
 QY 362 RFWNNTTNSHLSM--DTGSOVCNLVMSKNNELVSTHGYSONOIIWRYPTMSKLTATLGG 420
 DB 1087 RLMDCCGTGCVILGHSHNNRYSAIFSPN-GEIATCTSDOT-VKIMDQGGCKLKTILG 1144
 QY 421 HTYRYLYAISPDCQIYVAGDETLRFNMY 451
 DB 1145 HTNWFDIAFSPDGLKILASASHDQTVKIMDV 1175

RESULT 6
 HET1_PODAN STANDARD; PRT; 1356 AA.
 AC O00808;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Vegetatable incompatibility protein HET-E-1.
 GN HET-E1.
 OS Podospora anserina.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Lasiosphaeriales; Podospora.

NCBI_TaxID=5145;
 [1] SEQUENCE FROM N.A. MEDLINE=66008991; Pubmed=7557402;
 RA Saude S., Turco B., Bequeret J.;
 RT "A gene responsible for vegetative incompatibility in the fungus
 RT Podospira anserina encodes a protein with a GTP-binding motif and G
 RT beta homologous domain."
 RL Gene 162:135-139(1995).
 CC -1- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH
 CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,
 CC HET-C.
 CC -1- SIMILARITY: CONTAINS 10 WD REPEATS (TRP-ASP DOMAINS).
 CC -1- SIMILARITY: CONTAINS 1 NACHT DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L28125; AA85775.1; -
 DR HSSP: P04901; 1786;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 10.
 DR PRINTS: P000320; GPROTEINBRPT.
 DR SMART: SM00320; WD40; 10.
 DR PROSITE: PS50837; NACHT; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; 10.
 DR PROSITE: PS50082; WD_REPEATS_2; 10.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 DR GTP-binding; Repeat; WD repeat
 KW DOMAIN 294 629 NACHT
 FT NP_BIND 300 307 GTP (POTENTIAL).
 FT REPEAT 839 869 WD 1.
 FT REPEAT 881 911 WD 2.
 FT REPEAT 923 953 WD 3.
 FT REPEAT 965 995 WD 4.
 FT REPEAT 1007 1037 WD 5.
 FT REPEAT 1049 1079 WD 6.
 FT REPEAT 1091 1121 WD 7.
 FT REPEAT 1133 1163 WD 8.
 FT REPEAT 1175 1205 WD 9.
 FT REPEAT 1217 1247 WD 10.
 FT REPEAT 1259 1289 WD 11.
 FT REPEAT 1301 1331 WD 12.
 FT REPEAT 1356 1386 WD 13.
 FT REPEAT 1408 1438 WD 14.
 FT REPEAT 1460 1490 WD 15.
 FT REPEAT 1512 1542 WD 16.
 FT REPEAT 1564 1594 WD 17.
 FT REPEAT 1616 1646 WD 18.
 FT REPEAT 1668 1698 WD 19.
 FT REPEAT 1720 1750 WD 20.
 FT REPEAT 1772 1802 WD 21.
 FT REPEAT 1824 1854 WD 22.
 FT REPEAT 1876 1906 WD 23.
 FT REPEAT 1928 1958 WD 24.
 FT REPEAT 1980 2010 WD 25.
 FT REPEAT 2032 2062 WD 26.
 FT REPEAT 2084 2114 WD 27.
 FT REPEAT 2136 2166 WD 28.
 FT REPEAT 2188 2218 WD 29.
 FT REPEAT 2240 2270 WD 30.
 FT REPEAT 2292 2322 WD 31.
 FT REPEAT 2344 2374 WD 32.
 FT REPEAT 2396 2426 WD 33.
 FT REPEAT 2448 2478 WD 34.
 FT REPEAT 2500 2530 WD 35.
 FT REPEAT 2552 2582 WD 36.
 FT REPEAT 2604 2634 WD 37.
 FT REPEAT 2656 2686 WD 38.
 FT REPEAT 2708 2738 WD 39.
 FT REPEAT 2760 2790 WD 40.
 FT REPEAT 2812 2842 WD 41.
 FT REPEAT 2864 2894 WD 42.
 FT REPEAT 2916 2946 WD 43.
 FT REPEAT 2968 2998 WD 44.
 FT REPEAT 3020 3050 WD 45.
 FT REPEAT 3072 3102 WD 46.
 FT REPEAT 3124 3154 WD 47.
 FT REPEAT 3176 3206 WD 48.
 FT REPEAT 3228 3258 WD 49.
 FT REPEAT 3280 3310 WD 50.
 FT REPEAT 3332 3362 WD 51.
 FT REPEAT 3384 3414 WD 52.
 FT REPEAT 3436 3466 WD 53.
 FT REPEAT 3488 3518 WD 54.
 FT REPEAT 3540 3570 WD 55.
 FT REPEAT 3592 3622 WD 56.
 FT REPEAT 3644 3674 WD 57.
 FT REPEAT 3696 3726 WD 58.
 FT REPEAT 3748 3778 WD 59.
 FT REPEAT 3800 3830 WD 60.
 FT REPEAT 3852 3882 WD 61.
 FT REPEAT 3904 3934 WD 62.
 FT REPEAT 3956 3986 WD 63.
 FT REPEAT 4008 4038 WD 64.
 FT REPEAT 4060 4090 WD 65.
 FT REPEAT 4112 4142 WD 66.
 FT REPEAT 4164 4194 WD 67.
 FT REPEAT 4216 4246 WD 68.
 FT REPEAT 4268 4298 WD 69.
 FT REPEAT 4320 4350 WD 70.
 FT REPEAT 4372 4402 WD 71.
 FT REPEAT 4424 4454 WD 72.
 FT REPEAT 4476 4506 WD 73.
 FT REPEAT 4528 4558 WD 74.
 FT REPEAT 4580 4610 WD 75.
 FT REPEAT 4632 4662 WD 76.
 FT REPEAT 4684 4714 WD 77.
 FT REPEAT 4736 4766 WD 78.
 FT REPEAT 4788 4818 WD 79.
 FT REPEAT 4840 4870 WD 80.
 FT REPEAT 4892 4922 WD 81.
 FT REPEAT 4944 4974 WD 82.
 FT REPEAT 4996 5026 WD 83.
 FT REPEAT 5048 5078 WD 84.
 FT REPEAT 5100 5130 WD 85.
 FT REPEAT 5152 5182 WD 86.
 FT REPEAT 5204 5234 WD 87.
 FT REPEAT 5256 5286 WD 88.
 FT REPEAT 5308 5338 WD 89.
 FT REPEAT 5360 5390 WD 90.
 FT REPEAT 5412 5442 WD 91.
 FT REPEAT 5464 5494 WD 92.
 FT REPEAT 5516 5546 WD 93.
 FT REPEAT 5568 5598 WD 94.
 FT REPEAT 5620 5650 WD 95.
 FT REPEAT 5672 5702 WD 96.
 FT REPEAT 5724 5754 WD 97.
 FT REPEAT 5776 5806 WD 98.
 FT REPEAT 5828 5858 WD 99.
 FT REPEAT 5880 5910 WD 100.
 FT REPEAT 5932 5962 WD 101.
 FT REPEAT 5984 6014 WD 102.
 FT REPEAT 6036 6066 WD 103.
 FT REPEAT 6088 6118 WD 104.
 FT REPEAT 6140 6170 WD 105.
 FT REPEAT 6192 6222 WD 106.
 FT REPEAT 6244 6274 WD 107.
 FT REPEAT 6296 6326 WD 108.
 FT REPEAT 6348 6378 WD 109.
 FT REPEAT 6400 6430 WD 110.
 FT REPEAT 6452 6482 WD 111.
 FT REPEAT 6504 6534 WD 112.
 FT REPEAT 6556 6586 WD 113.
 FT REPEAT 6608 6638 WD 114.
 FT REPEAT 6660 6690 WD 115.
 FT REPEAT 6712 6742 WD 116.
 FT REPEAT 6764 6794 WD 117.
 FT REPEAT 6816 6846 WD 118.
 FT REPEAT 6868 6898 WD 119.
 FT REPEAT 6920 6950 WD 120.
 FT REPEAT 6972 7002 WD 121.
 FT REPEAT 7024 7054 WD 122.
 FT REPEAT 7076 7106 WD 123.
 FT REPEAT 7128 7158 WD 124.
 FT REPEAT 7180 7210 WD 125.
 FT REPEAT 7232 7262 WD 126.
 FT REPEAT 7284 7314 WD 127.
 FT REPEAT 7336 7366 WD 128.
 FT REPEAT 7388 7418 WD 129.
 FT REPEAT 7440 7470 WD 130.
 FT REPEAT 7492 7522 WD 131.
 FT REPEAT 7544 7574 WD 132.
 FT REPEAT 7596 7626 WD 133.
 FT REPEAT 7648 7678 WD 134.
 FT REPEAT 7700 7730 WD 135.
 FT REPEAT 7752 7782 WD 136.
 FT REPEAT 7804 7834 WD 137.
 FT REPEAT 7856 7886 WD 138.
 FT REPEAT 7908 7938 WD 139.
 FT REPEAT 7960 7990 WD 140.
 FT REPEAT 8012 8042 WD 141.
 FT REPEAT 8064 8094 WD 142.
 FT REPEAT 8116 8146 WD 143.
 FT REPEAT 8168 8198 WD 144.
 FT REPEAT 8220 8250 WD 145.
 FT REPEAT 8272 8302 WD 146.
 FT REPEAT 8324 8354 WD 147.
 FT REPEAT 8376 8406 WD 148.
 FT REPEAT 8428 8458 WD 149.
 FT REPEAT 8480 8510 WD 150.
 FT REPEAT 8532 8562 WD 151.
 FT REPEAT 8584 8614 WD 152.
 FT REPEAT 8636 8666 WD 153.
 FT REPEAT 8688 8718 WD 154.
 FT REPEAT 8740 8770 WD 155.
 FT REPEAT 8792 8822 WD 156.
 FT REPEAT 8844 8874 WD 157.
 FT REPEAT 8896 8926 WD 158.
 FT REPEAT 8948 8978 WD 159.
 FT REPEAT 9000 9030 WD 160.
 FT REPEAT 9052 9082 WD 161.
 FT REPEAT 9104 9134 WD 162.
 FT REPEAT 9156 9186 WD 163.
 FT REPEAT 9208 9238 WD 164.
 FT REPEAT 9260 9290 WD 165.
 FT REPEAT 9312 9342 WD 166.
 FT REPEAT 9364 9394 WD 167.
 FT REPEAT 9416 9446 WD 168.
 FT REPEAT 9468 9498 WD 169.
 FT REPEAT 9520 9550 WD 170.
 FT REPEAT 9572 9602 WD 171.
 FT REPEAT 9624 9654 WD 172.
 FT REPEAT 9676 9706 WD 173.
 FT REPEAT 9728 9758 WD 174.
 FT REPEAT 9780 9810 WD 175.
 FT REPEAT 9832 9862 WD 176.
 FT REPEAT 9884 9914 WD 177.
 FT REPEAT 9936 9966 WD 178.
 FT REPEAT 9988 10018 WD 179.
 FT REPEAT 10040 10070 WD 180.
 FT REPEAT 10092 10122 WD 181.
 FT REPEAT 10144 10174 WD 182.
 FT REPEAT 10196 10226 WD 183.
 FT REPEAT 10248 10278 WD 184.
 FT REPEAT 10300 10330 WD 185.
 FT REPEAT 10352 10382 WD 186.
 FT REPEAT 10404 10434 WD 187.
 FT REPEAT 10456 10486 WD 188.
 FT REPEAT 10508 10538 WD 189.
 FT REPEAT 10560 10590 WD 190.
 FT REPEAT 10612 10642 WD 191.
 FT REPEAT 10664 10694 WD 192.
 FT REPEAT 10716 10746 WD 193.
 FT REPEAT 10768 10798 WD 194.
 FT REPEAT 10820 10850 WD 195.
 FT REPEAT 10872 10902 WD 196.
 FT REPEAT 10924 10954 WD 197.
 FT REPEAT 10976 11006 WD 198.
 FT REPEAT 11028 11058 WD 199.
 FT REPEAT 11080 11110 WD 200.
 FT REPEAT 11132 11162 WD 201.
 FT REPEAT 11184 11214 WD 202.
 FT REPEAT 11236 11266 WD 203.
 FT REPEAT 11288 11318 WD 204.
 FT REPEAT 11340 11370 WD 205.
 FT REPEAT 11392 11422 WD 206.
 FT REPEAT 11444 11474 WD 207.
 FT REPEAT 11496 11526 WD 208.
 FT REPEAT 11548 11578 WD 209.
 FT REPEAT 11600 11630 WD 210.
 FT REPEAT 11652 11682 WD 211.
 FT REPEAT 11704 11734 WD 212.
 FT REPEAT 11756 11786 WD 213.
 FT REPEAT 11808 11838 WD 214.
 FT REPEAT 11860 11890 WD 215.
 FT REPEAT 11912 11942 WD 216.
 FT REPEAT 11964 11994 WD 217.
 FT REPEAT 12016 12046 WD 218.
 FT REPEAT 12068 12098 WD 219.
 FT REPEAT 12120 12150 WD 220.
 FT REPEAT 12172 12202 WD 221.
 FT REPEAT 12224 12254 WD 222.
 FT REPEAT 12276 12306 WD 223.
 FT REPEAT 12328 12358 WD 224.
 FT REPEAT 12380 12410 WD 225.
 FT REPEAT 12432 12462 WD 226.
 FT REPEAT 12484 12514 WD 227.
 FT REPEAT 12536 12566 WD 228.
 FT REPEAT 12588 12618 WD 229.
 FT REPEAT 12640 12670 WD 230.
 FT REPEAT 12692 12722 WD 231.
 FT REPEAT 12744 12774 WD 232.
 FT REPEAT 12796 12826 WD 233.
 FT REPEAT 12848 12878 WD 234.
 FT REPEAT 12900 12930 WD 235.
 FT REPEAT 12952 12982 WD 236.
 FT REPEAT 13004 13034 WD 237.
 FT REPEAT 13056 13086 WD 238.
 FT REPEAT 13108 13138 WD 239.
 FT REPEAT 13160 13190 WD 240.
 FT REPEAT 13212 13242 WD 241.
 FT REPEAT 13264 13294 WD 242.
 FT REPEAT 13316 13346 WD 243.
 FT REPEAT 13368 13398 WD 244.
 FT REPEAT 13420 13450 WD 245.
 FT REPEAT 13472 13502 WD 246.
 FT REPEAT 13524 13554 WD 247.
 FT REPEAT 13576 13606 WD 248.
 FT REPEAT 13628 13658 WD 249.
 FT REPEAT 13680 13710 WD 250.
 FT REPEAT 13732 13762 WD 251.
 FT REPEAT 13784 13814 WD 252.
 FT REPEAT 13836 13866 WD 253.
 FT REPEAT 13888 13918 WD 254.
 FT REPEAT 13940 13970 WD 255.
 FT REPEAT 13992 14022 WD 256.
 FT REPEAT 14044 14074 WD 257.
 FT REPEAT 14096 14126 WD 258.
 FT REPEAT 14148 14178 WD 259.
 FT REPEAT 14200 14230 WD 260.
 FT REPEAT 14252 14282 WD 261.
 FT REPEAT 14304 14334 WD 262.
 FT REPEAT 14356 14386 WD 263.
 FT REPEAT 14408 14438 WD 264.
 FT REPEAT 14460 14490 WD 265.
 FT REPEAT 14512 14542 WD 266.
 FT REPEAT 14564 14594 WD 267.
 FT REPEAT 14616 14646 WD 268.
 FT REPEAT 14668 14698 WD 269.
 FT REPEAT 14720 14750 WD 270.
 FT REPEAT 14772 14802 WD 271.
 FT REPEAT 14824 14854 WD 272.
 FT REPEAT 14876 14906 WD 273.
 FT REPEAT 14928 14958 WD 274.
 FT REPEAT 14980 15010 WD 275.
 FT REPEAT 15032 15062 WD 276.
 FT REPEAT 15084 15114 WD 277.
 FT REPEAT 15136 15166 WD 278.
 FT REPEAT 15188 15218 WD 279.
 FT REPEAT 15240 15270 WD 280.
 FT REPEAT 15292 15322 WD 281.
 FT REPEAT 15344 15374 WD 282.
 FT REPEAT 15396 15426 WD 283.
 FT REPEAT 15448 15478 WD 284.
 FT REPEAT 15500 15530 WD 285.
 FT REPEAT 15552 15582 WD 286.
 FT REPEAT 15604 15634 WD 287.
 FT REPEAT 15656 15686 WD 288.
 FT REPEAT 15708 15738 WD 289.
 FT REPEAT 15760 15790 WD 290.
 FT REPEAT 15812 15842 WD 291.
 FT REPEAT 15864 15894 WD 292.
 FT REPEAT 15916 15946 WD 293.
 FT REPEAT 15968 15998 WD 294.
 FT REPEAT 16020 16050 WD 295.
 FT REPEAT 16072 16102 WD 296.
 FT REPEAT 16124 16154 WD 297.
 FT REPEAT 16176 16206 WD 298.
 FT REPEAT 16228 16258 WD 299.
 FT REPEAT 16280 16310 WD 300.
 FT REPEAT 16332 16362 WD 301.
 FT REPEAT 16384 16414 WD 302.
 FT REPEAT 16436 16466 WD 303.
 FT REPEAT 16488 16518 WD 304.
 FT REPEAT 16540 16570 WD 305.
 FT REPEAT 16592 16622 WD 306.
 FT REPEAT 16644 16674 WD 307.
 FT REPEAT 16696 16726 WD 308.
 FT REPEAT 16748 16778 WD 309.
 FT REPEAT 16800 16830 WD 310.
 FT REPEAT 16852 16882 WD 311.
 FT REPEAT 16904 16934 WD 312.
 FT REPEAT 16956 16986 WD 313.
 FT REPEAT 17008 17038 WD 314.
 FT REPEAT 17060 17090 WD 315.
 FT REPEAT 17112 17142 WD 316.
 FT REPEAT 17164 17194 WD 317.
 FT REPEAT 17216 17246 WD 318.
 FT REPEAT 17268 17298 WD 319.
 FT REPEAT 17320 17350 WD 320.
 FT REPEAT 17372 17402 WD 321.
 FT REPEAT 17424 17454 WD 322.
 FT REPEAT 17476 17506 WD 323.
 FT REPEAT 17528 17558 WD 324.
 FT REPEAT 17580 17610 WD 325.
 FT REPEAT 17632 17662 WD 326.
 FT REPEAT 17684 17714 WD 327.
 FT REPEAT 17736 17766 WD 328.
 FT REPEAT 17788 17818 WD 329.
 FT REPEAT 17840 17870 WD 330.
 FT REPEAT 17892 17922 WD 331.
 FT REPEAT 17944 17974 WD 332.
 FT REPEAT 17996 18026 WD 333.
 FT REPEAT 18048 18078 WD 334.
 FT REPEAT 18100 18130 WD 335.
 FT REPEAT 18152 18182 WD 336.
 FT REPEAT 18204 18234 WD 337.
 FT REPEAT 18256 18286 WD 338.
 FT REPEAT 18308 18338 WD 339.
 FT REPEAT 18360 18390 WD 340.
 FT REPEAT 18412 18442 WD 341.
 FT REPEAT 18464 18494 WD 342.
 FT REPEAT 18516 18546 WD 343.
 FT REPEAT 18568 18598 WD 344.
 FT REPEAT 18620 18650 WD 345.
 FT REPEAT 18672 18702 WD 346.
 FT REPEAT 18724 18754 WD 347.
 FT REPEAT 18776 18806 WD 348.
 FT REPEAT 18828 18858 WD 349.
 FT REPEAT 18880 18910 WD 350.
 FT REPEAT 18932 18962 WD 351.
 FT REPEAT 18984 19014 WD 352.
 FT REPEAT 19036 19066 WD 353.
 FT REPEAT 19088 19118 WD 354.
 FT REPEAT 19140 19170 WD 355.
 FT REPEAT 19192 19222 WD 356.
 FT REPEAT 19244 19274 WD 357.
 FT REPEAT 19296 19326 WD 358.
 FT REPEAT 19348 19378 WD 359.
 FT REPEAT 19400 19430 WD 360.
 FT REPEAT 19452 19482 WD 361.
 FT REPEAT 19504 19534 WD 362.
 FT REPEAT 19556 19586 WD 363.
 FT REPEAT 19608 19638 WD 364.
 FT REPEAT 19660 19690 WD 365.
 FT REPEAT 19712 19742 WD 366.
 FT REPEAT 19764 19794 WD 367.
 FT REPEAT 19816 19846 WD 368.
 FT REPEAT 19868 19898 WD 369.
 FT REPEAT 19920 19950 WD 370.
 FT REPEAT 19972 20002 WD 371.
 FT REPEAT 20024 20054 WD 372.
 FT REPEAT 20076 20106 WD 373.
 FT REPEAT 20128 20158 WD 374.
 FT REPEAT 20180 20210 WD 375.
 FT REPEAT 20232 20262 WD 376.
 FT REPEAT 20284 20314 WD 377.
 FT REPEAT 20336 20366 WD 378.
 FT REPEAT 20388 20418 WD 379.
 FT REPEAT 20440 20470 WD 380.
 FT REPEAT 20492 20522 WD 381.
 FT REPEAT 20544 20574 WD 382.
 FT REPEAT 20596 20626 WD 383.
 FT REPEAT 20648 20678 WD 384.
 FT REPEAT 20700 20730 WD 385.
 FT REPEAT 20752 20782 WD 386.
 FT REPEAT 20804 20834 WD 387.
 FT REPEAT 20856 20886 WD 388.
 FT REPEAT 20908 20938 WD 389.
 FT REPEAT 20960 20990 WD 390.
 FT REPEAT 21012 21042 WD 391.
 FT REPEAT 21064 21094 WD 392.
 FT REPEAT 21116 21146 WD 393.
 FT REPEAT 21168 21198 WD 394.
 FT REPEAT 21220 21250 WD 395.
 FT REPEAT 21272 21302 WD 396.
 FT REPEAT 21324 21354 WD 397.
 FT REPEAT 21376 21406 WD 398.
 FT REPEAT 21428 21458 WD 399.
 FT REPEAT 21480 21510 WD 400.
 FT REPEAT 21532 21562 WD 401.
 FT REPEAT 21584 21614 WD 402.
 FT REPEAT 21636 21666 WD 403.
 FT REPEAT 21688 21718 WD 404.
 FT REPEAT 21740 21770 WD 405.
 FT REPEAT 21792 21822 WD 406.
 FT REPEAT 21844 21874 WD 407.
 FT REPEAT 21896 21926 WD 408.
 FT REPEAT 21948 21978 WD 409.
 FT REPEAT 22000 22030 WD 410.
 FT REPEAT 22052 22082 WD 411.
 FT REPEAT 22104 22134 WD 412.
 FT REPEAT 22156 22186 WD 413.
 FT REPEAT 22208 22238 WD 414.
 FT REPEAT 22260 22290 WD 415.
 FT REPEAT 22312 22342 WD 416.
 FT REPEAT 22364 22394 WD 417.
 FT REPEAT 22416 22446 WD 418.
 FT REPEAT 22468 22498 WD 419.
 FT REPEAT 22520 22550 WD 420.
 FT REPEAT 22572 22602 WD 421.
 FT REPEAT 22624 22654 WD 422.
 FT REPEAT 22676 22706 WD 423.
 FT REPEAT 22728 22758 WD 424.
 FT REPEAT 22780 22810 WD 425.
 FT REPEAT 22832 22862 WD 426.
 FT REPEAT 22884 22914 WD 427.
 FT REPEAT 22936 22966 WD 428.
 FT REPEAT 22988 23018 WD 429.
 FT REPEAT 23040 23070 WD 430.
 FT REPEAT 23092 23122 WD 431.
 FT REPEAT 23144 23174 WD 432.
 FT REPEAT 23196 23226 WD 433.
 FT REPEAT 23248 23278 WD 434.
 FT REPEAT 23300 23330 WD 435.
 FT REPEAT 23352 23382 WD 436.
 FT REPEAT 23404 23434 WD 437.
 FT REPEAT 23456 23486 WD 438.
 FT REPEAT 23508 23538 WD 439.
 FT REPEAT 23560 23590 WD 440.
 FT REPEAT 23612 23642 WD 441.
 FT REPEAT 23664 23694 WD 442.
 FT REPEAT 23716 23746 WD 443.
 FT REPEAT 23768 23798 WD 444.
 FT REPEAT 23820 23850 WD 445.
 FT REPEAT 23872 23902 WD 446.
 FT REPEAT 23924 23954 WD 447.
 FT REPEAT 23976 24006 WD 448.
 FT REPEAT 24028 24058 WD 449.
 FT REPEAT 24080 24110 WD 450.
 FT REPEAT 24132 24162 WD 451.
 FT REPEAT 24184 24214 WD 452.
 FT REPEAT 24236 24266 WD 453.
 FT REPEAT 24288 24318 WD 454.
 FT REPEAT 24340 24370 WD 455.
 FT REPEAT 24392 24422 WD 456.
 FT REPEAT 24444 24474 WD 457.
 FT REPEAT 24496 24526 WD 458.
 FT REPEAT 24548 24578 WD 459.
 FT REPEAT 24600 24630 WD 460.
 FT REPEAT 24652 24682 WD 461.
 FT REPEAT 24704 24734 WD 462.
 FT REPEAT 24756 24786 WD 463.
 FT REPEAT 24808 24838 WD 464.
 FT REPEAT 24860 24890 WD 465.
 FT REPEAT 24912 24942 WD 466.
 FT REPEAT 24964 24994 WD 467.
 FT REPEAT 25016 25046 WD 468.
 FT REPEAT 25068 25098 WD 469.
 FT REPEAT 25120 25150 WD 470.
 FT REPEAT 25172 25202 WD 471.
 FT REPEAT 25224 25254 WD 472.
 FT REPEAT 25276 25306 WD 473.
 FT REPEAT 25328 25358 WD 474.
 FT REPEAT 25380 25410 WD 475.
 FT REPEAT 25432 25462 WD 476.
 FT REPEAT 25484 25514 WD 477.
 FT REPEAT 25536 25566 WD 478.
 FT REPEAT 25588 25618 WD 479.
 FT REPEAT 2564

Db 265 -----GEYQREGQIAPVSD-----TEAARTTSQSWYVYINPCKRV 301
 QY 169 FYLNLVDWSSNNVLAVGLGNVYLMNACSSKVTKLCDLGVDDVCVGAAGGTFLAVGT 228
 Db 302 FMINVHTLH-----PSYCCVCKSSNNGKTLATGC 332
 QY 229 NNGKQVQWDAARCKIRSM-----GHLRLVGLAMS-----SSLSGGGDKNTYQDRT 280
 Db 333 NQA-ANVEDVQTKLFTLHEESPDSRDLVYRTAFSDGKYLVLTGEDRQIKMLDST 391
 QY 281 QE-DEVSRLSGHSEVQGLKMSVDNRELASGDNKLFVWQHSOTQPLKYCEHTAAYKA 339
 Db 392 QKAYVY--FSGHEODIYSLDFSHNGRFYVSGSDRTARLMPVETGQCILK-LEIENGYTA 448
 QY 340 IAMSPLHGLASGGGTADRCIRFNNNTTNSHLSGMDTGSQVQCNLMVSKNNELVSTHY 399
 Db 449 IALSPMDQFLAV---GSLDOIIRVWSVSGTLVERLEGHKESVYIASSPSSILLS--GS 503
 QY 400 SONQIIVWRYPTMSKL-----ATLGHTRYLVLAISPDGQITVTCAGDETLR 447
 Db 504 LKTIKVMELQATRSVGLSAIKPEGICAKATYTGHTFVLVAVSPDSRMGLSGSKDSMQ 563
 QY 448 FNNV 451
 Db 564 FMDL 567

RESULT 8

TU12_SCHPO STANDARD; PRT; 586 AA.
 AC 09UCG8: 014432: (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcriptional repressor tup12.
 GN TU12 OR TU01 OR SPAC630.14C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetes; Schizosaccharomycetaceae;
 OC NCBI_Taxid=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA McDougall R.C., Rajandream M.A., Barrell B.G., Zimmermann W.,
 RA Wambolt R.;
 RN Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE OF 26-586 FROM N.A.
 RC STRAIN=972;
 RA Braun B.R.;
 RN Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RP CHARACTERIZATION.
 RX MEDLINE=20036803; Pubmed=10567571;
 RA Mukai Y., Matsuo E., Roth S.Y., Harashima S.;
 RT "Conservation of histone binding and transcriptional repressor
 RT Mol. Cell. Biol. 19:8461-8468(1999)."
 CC - FUNCTION: TRANSCRIPTIONAL REPRESSOR.
 CC - SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC - SIMILARITY: BELONGS TO THE TU01 FAMILY OF WD-REPEAT PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC entitled and this statement is not removed. Usage by and for commercial
 CC or send an email to license@ebi-sib.ch).
 DR EMBL: AL109832; CAB52736.1;
 DR EMBL: U92792; AAB81475.2;

DR InterPro: IPR001632; Gprotein_B.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR PRINTS: PRO0319; GPROTEINB.
 DR PRINTS: PRO0320; GPROTEINBPT.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS00678; WD_REPEATS_1; 3.
 DR PROSITE: PS50082; WD_REPEATS_2; 6.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Transcription regulation; Repressor; Repeat; WD repeat.
 FT REPEAT 325
 FT REPEAT 332
 FT REPEAT 371
 FT REPEAT 374
 FT REPEAT 413
 FT REPEAT 415
 FT REPEAT 456
 FT REPEAT 495
 FT REPEAT 510
 FT REPEAT 549
 FT REPEAT 552
 FT REPEAT 585
 FT REPEAT 64192 MW; 79C991ABDDC7D095 CRC64;
 SEQUENCE 586 AA; 64192 MW; 79C991ABDDC7D095 CRC64;

Query Match

Best Local Similarity 9.8%; Score 250; DB 1; Length 586;
 Matches 117; Conservative 73; Mismatches 175; Indels 146; Gaps 23;

QY 5 GNRNPPPTSTVRDNDSPPP-----EPSPESLR-----HVSRY-INSNH 41
 Db 130 GNAGPFNS-----DNSSKPLILNNGSSGTFNNLNSPAIDSDGTVLAPIQTSNVDLSQYY 185
 QY 42 TSP-----SRITYSD--RFIPSNASKFALFDINPTREGDDSSAVYTLRLALFGPD 93
 Db 186 SSPHRAVAGATMGASAMRTFSPN-----LPLGHPPPPSDSANSVPPIAPLVNKG 238
 QY 94 VAG-VPTPEKTDSPSMTLPNNIRPKYKTRQSMHSLSPMDDEYVGVNHSVPVAPRKV 152
 Db 239 VSGNPPYPAEI-IFTSNVPRE-----EKDMTVYSN-----VPKKE 273
 QY 153 PRSPKVLVDAPODDPYLLVLYOMSSH-NVLAVGLGNCYVLMNACSSKVTKLC----- 204
 Db 274 P-PLVQLHTLHTHSVICYVRFESADGKFLATGCRAMAVENVETGKILITLLODESSKR 331
 QY 205 --DLGVDDCVCSYGMARGTHTLAVGTNNKGVOIYMAARCKIRSEGHRLVGLAMSSS 262
 Db 332 EGDL---YVRSVAFSPDGKYLATGVEDQIRINDIAO----- 365
 QY 263 LLSGGGDKNTYQDRTIRQEDVSKLSGHKSEVCGLKMSYDNRELASGDNKLFVWQHS 322
 Db 366 -----KRYR-----LITGHEQEIYSLDFSKDKGLTVSGSDRTVCLMDVE 406
 QY 323 STQPVLYKCEHT-AAVKALAMSPHLHGLASGGGTADRCIRFNNNTTNSHLSGMDTGSQV 381
 Db 407 AGEQ--KLILHTDDGVTVTFMSPDQFLIA---GSLDKVIRIWTSSQTLVEQLHGEESV 461
 QY 382 CNLWMSKNNELVSTHSGSONQIIVWRYPTMSKLA-----TLVGHTRYLVLYA 429
 Db 462 YSVASPPQKLYV--GSLDNTIKLMELQCVSNVASMKEGICQOTFTGHKDPILSVT 519
 QY 430 ISPDQITVTAAGDETLRFNWNVFPSPQSNQ 460
 Db 520 VSPDGKWLISGSKDRITQFWS-DSPHSQLT 549
 RESULT 9
 YCW2_YEAST STANDARD; PRT; 515 AA.
 AC P25382;
 DT 01-MAV-1992 (Rel. 22, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypochemical 57.0 kDa trp asp repeats containing protein in CPR4-SSK22
 DE intergenic region.
 GN YCRO72C OR YCR72C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

10

100

QY 151 -KVRSPYKVLADAPALODDFYLNLDKSSHNVLAVGLGCVLMACSSKYTKLDLGYD 209
 DB 172 EKKOHEEYVLYNPAFSKDLIDMHSLSHS----- 202
 QY 210 DVCVSGVAGRGTHLAAGTNNKVQIWDARCKIRSM-----EGRRLRVG 255
 DB 203 SVCCVAFSFDGKFIATGCKR-TTOVFNTTGTGLVAKLIDSSSENKDDNTTASGDLTYR 261
 QY 256 ALAMS--SLLSSGGRDNKIYORDIRTOEDDFYSLGSKSVGCGKMSYDNRELASGND 313
 DB 262 SVCFSPDGKRLTAGADKILIRIMDLSTR-IKILRGHEDDITSLDFPDPGDRLYVSGSGD 320
 QY 314 NKLEFVNOHSTOPLVLYKCEHTAAVK-----ALAMSPHLGLLASGCGTADRCIRFNNWTTN 369
 DB 321 RSVRIWDLRTSQ-----CSLTLIEDGVTVAVSP--DGKILR-AGSLDRIVKLMHLEGRSKSCEV 430
 QY 370 SHLSCMDTSSQ-----VCNLVMSKNVNELVSTHGSQNOIIVRYTMSKL-----A 416
 DB 373 FVEVRLDSNGENGHEDSVYSAFNSNGEQIAS--GSLDRIVKLMHLEGRSKSCEV 430
 QY 417 TLTGHTYVLYLAISPDGQITVAGDETLRFNN 450
 DB 431 TYIGHKDFVLSVCTCPDNEIILSGSKDRGVIFWD 464

RESULT 11

RCOL_NEUCR STANDARD: PRT: 604 AA.
 AC P78706: 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Transcriptional repressor rco-1.
 GN RCO-1.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariaceae; Sordariaceae; Neurospora.
 RN NCBI_TaxID=5141;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97042452; PubMed-8887652;
 RA Yamashiro C.T., Ebdole D.J., Lee B.-U., Brown R.E., Bourland C.,
 RT Madl L., Yanofsky C.;
 RT "Characterization of rco-1 of Neurospora crassa, a pleiotropic gene
 affecting growth and development that encodes a homolog of Tup1 of
 Saccharomyces cerevisiae.";
 RL Mol. Cell. Biol. 16:6218-6228(1996).
 CC -1- FUNCTION: REPRESSES TRANSCRIPTION BY RNA POLYMERASE II. MAY BE
 INVOLVED AT SEVERAL STAGES OF CONDITATION AND OTHER GROWTH AND
 DEVELOPMENT PROCESSES. APPEARS TO REGULATE GENES THAT ARE
 EXPRESSED IN ASEYUAL AND SEXUAL SPORE PATHWAYS. MUTANT RCO-1
 ALLELES ARE ACONIDIAL, FEMALE STERILE, HAVE REDUCED GROWTH RATES
 AND FORM HYPAE THAT COIL IN A COUNTERCLOCKWISE DIRECTION,
 OPPOSITE TO THAT OF THE WILD TYPE.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL: U57061; AAB37245.1;
 DR Interpro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR PRINTS: PRO0320; GPROTEINBRPT.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS00678; WD_REPEATS_1; 4.
 DR PROSITE: PS50082; WD_REPEATS_2; 6.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Transcription regulation; Repressor; Repeat; WD repeat.

FT REPEAT 295 324 WD 1.
 FT REPEAT 342 372 WD 2.
 FT REPEAT 384 414 WD 3.
 FT REPEAT 425 455 WD 4.
 FT REPEAT 469 499 WD 5.
 FT REPEAT 523 553 WD 6.
 FT REPEAT 565 600 WD 7.
 FT DOMAIN 130 135 POLY-PRO.
 FT DOMAIN 153 160 POLY-PRO.
 FT DOMAIN 244 249 POLY-PRO.
 SQ SEQUENCE 604 AA; 66076 MW; C962B5E558B529A4 CRC64;

Query Match Best Local Similarity 8.5%; Score 217; DB 1; Length 604;
 Matches 109; Conservative 84; Mismatches 202; Indels 136; Gaps 24;

QY 5 GNRNPPPTSTVIRNSP-----PPSPESLRHVSRIANSNHYT 42
 DB 92 GNMHPFHGQOOPPAIGSNVFSALIMAGGQALVPPPPPOOOPAHNPAPGLQ 151
 QY 43 SP-----SRTIYDRFIPERSASRFALEPDITPEEGDDSSAYTLLFTALFPPV 94
 DB 152 GPPPPPPPSQPPFQOQYQGGQGNF-----PPQPSYASGPAKR-GIGRPA 203
 QY 95 AGVTPPE-----KTDSPM-TLNNRIEYKTEKTRSHSLSPMDDPFGVNH 143
 DB 204 GGPATPQINTPIFYNGGASQPVPTHTPD-----HTRMQH-----H 242
 QY 144 SPVKAARKVPSPKYKVLADAPALODDFYLNLDKSSHNVLAVGLGCVLMACSSKYTKL 203
 DB 243 QPPPPPS-QTNALSELDPRLPHIKKKMDK-----WVIFNAAPRV 285
 QY 204 CDLGV-----DPCVSGVAGRGTHLAAGTNNKVQIWDARCKIRSMGHR----- 251
 DB 286 LDVELVHTLHESVYCCVRSMDGKIVAVGCNR-SAQIVDETGEKCLILQDEIDITGD 344
 QY 252 LRVALGAS--SLLSSGGRDNKIYORDIRTOEDFVSLSKSEVGLMSTYDNRELAS 309
 DB 345 LYRSVCFSPDGKRLTAGADKILIRIMDIQSR-TIRMTFCHEDDITSLDFSGDRTIAS 403
 QY 310 GGNKNTLVNNOHSTO--PVLKCEHTAAVKALAMSPHLGLLASGCGTADRCIRFNNWTT 367
 DB 404 GSGDRIVRLMDETGCQNNVSLIED--GVTVAISDQKFVAA--GSLDSVAVMD-- 455
 QY 368 TNSHSCMDT-----SOYCNLWMSKNVNELVSTHGSQNOIIVR-----YPMRK-- 414
 DB 456 MGVIALERLEGPBGDKDSVSAFSPDGRNLVS--GSLDKTKMELSLAPKQIPSPAPK 513
 QY 415 ----LATLTGHTYVLYLAISPDGQITVAGDETLRFNNVFPSPKSONTE 461
 DB 514 GGRCTKTEGHRDFVLSVALTPPSQWVLSGSKDRGVIFWD-----PRTGHTQ 560

RESULT 12

PRLL_ARATH STANDARD: PRT: 486 AA.
 AC 042384: 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE PPL1/PP2A phosphatases pleiotropic regulator PRL1.
 GN PRL1 OR AT4G15900 OR PL3990W.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Nemeth K., Putnocky P., Stankovic B., Bako L., Mathur J., Redel G.P.,
 RA Schnell J., Konc C.,

to the EMBL/GenBank/DBJ databases

PROSITE: P550294; WD_REPEATS_REGION: 1.	Score 209; DB 1; Length 486;
RM Repeat: 174 204 WD 1.	
FT Repeat: 216 246 WD 2.	
FT Repeat: 258 288 WD 3.	
FT Repeat: 300 330 WD 4.	
FT Repeat: 342 371 WD 5.	
FT Repeat: 384 413 WD 6.	
FT Repeat: 433 463 WD 7.	
SEQUENCE 486 AA; 54009 MW; 604AD3E6FCDA8A998 CRC64;	
Query Match: 8.2%; Score 209; DB 1; Length 486;	
Query Local Similarity: 21.0%; Pred No. 5,1e-09;	
Matches 103; Conservative 78; Mismatches 197; Indels 112; Gaps 17	
13 STYRNRNSPPPEPSPESTRHVSRM-----INSNHTSPRTIYSRPIFSRSASKFA 63	
32 SPHGCGPPPEPEAKQILSHKMKVAFGVEPVASQPPROPDR--INEGPSSNLSIAA 89	
64 LFDINTEGRRDSSATITLRTALRGPDVAGPVPE-----KIDSPTMLPNRIFRY 118	
90 PEDSKSTOGAGESA-----IVGPTLLRLPLKGLNTGSSGKSTIIIPANVSSY 140	
119 KTEROSMHSLSPEFMDDFVPGVNHSPYKARRYPRSPYKVLDAFALODEFVILNVDS 178	
141 -----QNNISALAME-----RISRPFRPW--NAP-----WRN 168	
179 HNVLAVALG-----NCVYLINACSSVYTKLCGLGVDDCVCSYGMA 218	
169 YRVIOGHGVWVRSAFPEPSNEMFCTGSADRTIKINDVAFG-VKLITLGHIEGRGLAVS 227	
219 ORGTHLAVGTNNKVOYIWDAAKCKTIRSMEGHRLRYGALAMSSSL--LSGGRKNLYOR 276	
228 NHTMTMFAGSDCKDYKCMDEQNKVIRSYGHLSGVYCLALHPTLDVLLTGRDSYCRW 287	
277 DIRTODEPVSKISGKSEVCGLKWSYNRELASGGDNKLFVWNOHSTQPVLYKTCERTAA 336	
288 DIRTQKQ-IFALSGHDNTVCSVFTRPDPVGVGSHDTTKFMDLRYGKTMSTLTHNRKS 346	
337 VKAIAMSPHLHGLLASGGGTADRCIRFENNTTNSHLSMDTGSQVONLVMSKNVE-LVS 395	
347 VRAMTLLHPEN--AFASASANTKKF--SLPGECHNMLSOQKTIINAMVNDGVW 400	
396 THGYSQNOIIVWRYPTMSKLAITLGHTRYLYLAISSP-----DQGIYTF 439	
401 VTGGDNGSIFWPMK-----SGHSFOOSTIIVQPSLESEAGIYACDYGRSLVT 452	
440 GAGETLRFW 449	
453 CEADKTIKMW 462	
RESULT 13	
CIAL HUMAN STANDARD; PRT: 339 AA.	
ID CIAL HUMAN	
AC 16-OCT-2001 (Rel. 40, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	
DE MD40-repeat containing protein clao 1.	
GN CIAO1.	
OS Homo sapiens (Human)	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
OC NCBI_Taxid=9606;	
RA MEDLINE=88225157; PubMed=9556563;	
RT Johnstone R.W., Wang J., Tommerup N., Vissing H., Roberts T., Shi Y.;	
RT "Clao 1 is a novel WD40 protein that interacts with the tumor	
RT suppressor protein Wt1.";	
RT J. Biol. Chem. 273:10880-10887(1998).	

	RESULT	4	PRIEST	
ID	PRO4_YEAST			
AC	P20053.	STANDARD:	PRF;	465 AA.
DT	01-FEB-1991 (Rel., 17, Created)			
DT	01-FEB-1991 (Rel., 17, Last sequence update)			
DT	15-JUL-1998 (Rel., 36, last annotation update)			
DN	U4/U6 small nuclear ribonucleoprotein PRP4			
GK	PPR4 OR RNA4 OR YPR178W OR p9705.6.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes.			
CC	NBLTaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89384596; PubMed=2528666;			
RA	Petersen-Bjorn S., Soltyk A., Beggs J.D., Friesen J.D.;			
RT	"PRP4 (RNA4) from Saccharomyces cerevisiae: its gene product is			
RL	associated with the U4/U5 small nuclear ribonucleoprotein particle."			
RM	Mol. Cell. Biol. 9:3698-3709(1989).			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89384597; PubMed=2528687;			
RA	Barniques J., Abelson J.N.;			
RT	"PRP4: A protein of the yeast U4/U6 small nuclear ribonucleoprotein			
RL	particle." [3]			
RM	Mol. Cell. Biol. 9:3710-3719(1989).			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-S286C / AB972;			
RA	Johanson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,			
RA	Favellio A., Fulton L., Gattling S., Greco T., Kirsten J., Kucaba T.,			
RA	Helmsthorff K., Hawkins J., Hillier L., Jler M., Johnson D.,			
RA	Johnston I., Langston Y., Letralle P., Le T., Merdis E., Meneses S.,			
RA	Taylor N., Nhan M., Pauley A., Peluso D., Rifken L., Riley L.,			
RA	Wilson R., Watskins E., Vignati D., Wilcox L., Wohlmann P., Vaudin M.,			
RL	Submitted (APR-1995) to the EMBL/GenBank/DDJ databases.			
-I-	FUNCTION: INVOLVED IN RNA SPLICING. IS REQUIRED FOR THE ASSOCIATION OF U4/U6 SNRNP WITH U5 SNRNP IN AN EARLY STEP OF SPICIOSOME ASSEMBLY.			
CC	-I- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).			
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its contents are in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-stb.ch/announce/ or send an email to license@isb-stb.ch). -----			
DR	EMBL_M26597; AAA79332.1; .			
DR	EMBL_M28518; AAA79011.1; .			
DR	EMBLL_025842; AAB68111.1; .			
DR	PIR_A32569; A32569.			
DR	SGD_S0006382; PRP4.			
DR	InterPro: IPRO035648; SEM.			
DR	pfam: PF001680; WD40.			
DR	PRINTS_PRO0320; GPOTEINBRPT.			
DR	SMART_SM00500; SEM; 1.			
DR	SMART_SM00320; WD40; 7.			
DR	PROSITE_PS00678; WD_REPEATS_1; 3.			
DR	PROSITE_PS50082; WD_REPEATS_2; 4.			
DR	mRNA_processing_Repeatc_WD_REPEATS_REGION; 1.			
KT	REPEAT 216 247			
FT	REPEAT 263 293			
FT	REPEAT 305 335			
FT	REPEAT 347 377			
FT	REPEAT 391 423			
FT	REPEAT 435 465			
WT	WD 5.			

SO SEQUENCE 465 AA: 52442 MW: 70CCCA3A2AC0CE78 CRC64:

Query Match 7.9%: Score 203; DB 1; Length 465;
Best Local Similarity 22.1%: Pred. No. 1.4e-08;
Matches 70; Conservative 52; Mismatches 107; Indels 88; Gaps 12;

197 SSKVTKLCLDGLVDDCVCSVG-WAORGTILAVGNNCKVQIMDAARCKKI-RSMGHRRLV 254
174 STKPLSAVSLSDMDVMVATGSGMA-----GDLYVNSQTLQPLTKLDSHWGKI 221
255 GALAW-----SSSLLS 265
222 GADLHMPDSNNOMISCADGLIKNFQY SNEGGILRLGLVGHERRISVYKHPGKRTIG 281
266 SGGGRDNIVQRIQEDPFVSKLSGKSEYQGLKSYDNRELASGGNDKLFVWNOHSTQ 325
282 SASHDMTWRLMDASTHOELLQ-EGHDKGVSLSFQCDGSLVSCGMDLSMLMDIRSGS 340
326 PVLKCEHTAAVKAIAWSPHLGLASGGGTADRCIRFNMNTTNSHSCMDTGSQVCLV 385
341 KVMPLASHKPIYTVAMSPNGY-QVATGGG--DGLINWMD-----IRKRDG-QLNQTL 390
386 WSKNV-----NELVTHGYSOMQIIVWRYPTMSKLTATLTGTYRVLALSPD 433
391 AHRNIVQVRFSEKEDGKGLVSC-GY-DNLINVSSTPLWLMKSLAGHTDKTILDISIN 448
434 GQTYTGAGDETLRFMN 450
449 SHELVSGGMDRSIKLMN 465

RESULT 15
YG4S-YEAST STANDARD; PRT: 409 AA.
ID YG4S-YEAST

AC 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 46.7 kDa protein in PER54-DIE2 intergenic region.
GN YGR225W OR G8541.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NC NCBI_TaxID=4932;
RM [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=96267763; PubMed=8701610;
RA van der Aart Q.J.M., Kleine K., Steensma H.Y.;
RT "Sequence analysis of the 4.3 kb CRM1-YLM9-PET54-DIE2-SM11-PHO81-YH84-
RT PFK1 region from the right arm of Saccharomyces cerevisiae chromosome
RT VII.";
RT Yeast 12:385-390(1996).
RL -1- SIMILARITY: SOME, TO YEAST CDC20 AND S.POMBE SPAC1366.08.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X87941; CAA61174.1;
DR EMBL: Z73010; CAA97253.1;
DR SGD: S0003457; SP070.
DR InterPro: IPR000002; FIZZY.
DR ProDom: PD004563; FIZZY.1.
KW Hypothetical protein.
SQ SEQUENCE 409 AA: 46660 MW: 91F7A246A2892AD6 CRC64;

Query Match 7.8%: Score 200; DB 1; Length 409;
Best Local Similarity 26.2%: Pred. No. 2.1e-08;
Matches 71; Conservative 45; Mismatches 85; Indels 70; Gaps 11;

13 STVRDNPPE--PSPESLRHVSRLMNSNHTSPSRITYSDRFLPSRSKFAIPDINTP 70
55 SELEKSPSPERLSPD-----FNEELRNTGHEHESIST-----NEFSMSSIS- 98
71 TEGRDSSSAVTLTLRTALEGPDVAGVPTPEKTD-----SPSMTLPNNIRRYKTEPR 123
99 -----SSSQVYRSGSARASRNDYSLKLEQDKRKNIAHSLGFLPDR-VFTEETTS 151
124 QSMH-----SLSPFMDDEFPVGVNHSPPVAKRPKVRP- 155
152 EILEKNKAIKNCFGPGSCAEIRSTFESTLSPVARYIANSNARSASPOQIOIRPAKRV 211
156 ---PYKVLDPALODPEYLVLDVMS--SHNVLAVALGNCVLYLNACSKYTKLCLDGLVD 209
212 KSHIPYRVLDAFCRLNDPYSMLISMSRTNNVL-VGIGGCVYIWSSEKAGAVSILDHOYLS 270
210 ---DCVCSVGMARGTILAVGTNNKQVQIMD 237
271 EKRDLVTCVSPCPYNTYFVGTGKFRILLID 301

Search completed: May 21, 2002, 11:50:25
Job time: 213 sec

Wed May 22 11:09:50 2002

us-09-701-572-2.rsp

Wed May 22 11:09:51 2002

us-09-701-572-2.rspt

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 11:46:22 ; Search time 31.11 Seconds
(without alignments)
2641.356 Million cell updates/sec

Title: US-09-701-572-2
Perfect score: 2555
Sequence: 1 MDGTGNRPPTSTVRDNPSP.....KQNTFSEIGALSGRTIR 475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_PRODUCE:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2555	100.0	475	10	Q9XED5 medicago va
2	2547	99.7	475	10	Q9W712 medicago va
3	1839	72.0	444	10	Q82740 arabidopsis
4	1825	71.4	482	10	Q9T060 arabidopsis
5	1448	56.7	472	10	Q9PFY8 arabidopsis
6	1294	50.6	493	4	Q7S869 homo sapien
7	1294	50.6	504	4	Q9UJH8 homo sapien
8	1286.5	50.4	496	4	Q9UM11 homo sapien
9	1286	50.3	493	11	Q9RIK5 mus musculu
10	1285	50.3	493	13	Q42585 xenopus lae
11	1276	48.9	478	5	Q18402 drosophila
12	1276	48.9	478	5	Q9W4H9 drosophila
13	1274	45.9	493	4	Q9UJ96 homo sapien
14	1170.5	45.8	528	10	Q94DF9 oryza sativ
15	1096	42.9	702	5	Q09649 caenorhabd
16	1066	41.7	556	3	Q13286 schizosacch

RESULT	1	PRELIMINARY	PRT	475 AA.	
Q9XED5	Q9XED5	Q9XED5	Q9XED5	Q9XED5	Q9XED5
AC	01-NOV-1999	(TREMBLrel. 12, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	CELL CYCLE SWITCH PROTEIN.				
GN	CCS52.				
OS	Medicago varia.				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
CC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Trifoliales; Medicago.				
OX	NCBI_TaxID=36902;				
RM	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=NODE;				
RA	Cebolla A., Kondorosi E.;				
RT	"A plant cell cycle regulator that controls endoreduplication and cell				
RT	differentiation."				
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: CONTAINS 6 WD REPEATS (TIR-ASP DOMAINS).				
DR	EMBL: AF079404; AAD22612.1;				
DR	InterPro: IPR000002; F1Zy.				
DR	InterPro: IPR001680; WD40.				
DR	Pfam: PF00400; WD40_5.				
DR	PRINTS: PR00320; GPROTEINRPT.				
DR	PRODOM: PD004563; F1Zy; 1.				
DR	SMART: SM00320; WD40_5.				
DR	PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.				
DR	PROSITE: PS00682; WD_REPEATS_2; 3.				
DR	PROSITE: PS0294; WD_REPEATS_REGION; 1.				
KW	Repeat; WD repeat.				
SQ	SEQUENCE 475 AA; 52542 MW; 8E54F0DAFAC018FB CRC64;				

Query Match 100.0%; Score 2555; DB 10; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.2e-211;
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDGTGNRPPTSTYRONSPPPEPSPESLRHSVMINSNHYTSPSTIYSDRFIPRSAS 60
 DB 1 MDGTGNRPPTSTYRONSPPPEPSPESLRHSVMINSNHYTSPSTIYSDRFIPRSAS 60
 QY 61 KPALFDINTPTGEGDDSSAYTTLLRTALFGPDVAGPYVPEKTDSPKMLPNNIFRYKT 120
 DB 61 KPALFDINTPTGEGDDSSAYTTLLRTALFGPDVAGPYVPEKTDSPKMLPNNIFRYKT 120
 QY 121 ETRQSMHSLSPFMDDDVPVGNHSPKAPRKPRSPYKVLADAPALODDFYLNLDVMSHN 180
 DB 121 ETRQSMHSLSPFMDDDVPVGNHSPKAPRKPRSPYKVLADAPALODDFYLNLDVMSHN 180
 QY 181 VLAAGLNCYVLLWNAACSSKVTKLCDLGVDDCVCSYGMORGHLAGVNNKVOIWDAR 240
 DB 181 VLAAGLNCYVLLWNAACSSKVTKLCDLGVDDCVCSYGMORGHLAGVNNKVOIWDAR 240
 QY 241 CKKIRSMEGHRLRVGLAWSSSLSSGGDKNITQDRIETQEDFVSKLSGHSKSEVCLKW 300
 DB 241 CKKIRSMEGHRLRVGLAWSSSLSSGGDKNITQDRIETQEDFVSKLSGHSKSEVCLKW 300
 QY 301 SYDNRELASGGDNKLFVWNOHSTQPVLYKCEHTAAVKAIAWSPHLGLASGGTADRC 360
 DB 301 SYDNRELASGGDNKLFVWNOHSTQPVLYKCEHTAAVKAIAWSPHLGLASGGTADRC 360
 QY 361 IREWNTTNSHLSGMDTSGVNCNLWMSKNVNLVSTHGYSONOIIYWRYPYMSKLTATL 420
 DB 361 IREWNTTNSHLSGMDTSGVNCNLWMSKNVNLVSTHGYSONOIIYWRYPYMSKLTATL 420
 QY 421 HTYRVLYLAISPDDGQITVTGAGDETLRFVWVFPSPKSQNTSEIGALSIGRTTIR 475
 DB 421 HTYRVLYLAISPDDGQITVTGAGDETLRFVWVFPSPKSQNTSEIGALSIGRTTIR 475

RESULT 2
 Q9M712 PRELIMINARY: PRT: 475 AA.
 AC Q9M712;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
 GN WM-REPEAT CELL CYCLE REGULATORY PROTEIN.
 OS Medicago truncatula (Barrel medic).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
 RN [1]
 NCBI_TaxID=3880;
 RP SEQUENCE FROM N.A.
 RC STRAIN=GHOR;
 RA Cebolla A., Vinardell J.M., Olah B., Kiss E., Roudier F.,
 RA Kondorosi A., Kondorosi E.;
 RT "The mitotic inhibitor ccs52 is required for endoreplication and
 RT polyploid-dependent cell enlargement in plants."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AF134835; AAF37386.1;
 DR InterPro: IPR000002; Fizzzy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.6.
 DR PRINTS: PRO0320; GPROTEINBRPT.
 DR PRODOM: PD004563; Fizzzy.1.
 DR SMART: SM00320; WD40.5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
 DR PROSITE: PS50294; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SO SEQUENCE 475 AA; 52457 MW; 2778EB31FD20D5C2 CRC64;

Query Match 99.7%; Score 2547; DB 10; Length 475;
 Best Local Similarity 99.6%; Pred. No. 5.8e-211;

Matches 473; Conservative 1; Mismatches 1; Indels 0; Gaps 0.
 QY 1 MDGTGNRPPTSTYRONSPPPEPSPESLRHSVMINSNHYTSPSTIYSDRFIPRSAS 60
 DB 1 MDGTGNRPPTSTYRONSPPPEPSPESLRHSVMINSNHYTSPSTIYSDRFIPRSAS 60
 QY 61 KPALFDINTPTGEGDDSSAYTTLLRTALFGPDVAGPYVPEKTDSPKMLPNNIFRYKT 120
 DB 61 KPALFDINTPTGEGDDSSAYTTLLRTALFGPDVAGPYVPEKTDSPKMLPNNIFRYKT 120
 QY 121 ETRQSMHSLSPFMDDDVPVGNHSPKAPRKPRSPYKVLADAPALODDFYLNLDVMSHN 180
 DB 121 ETRQSMHSLSPFMDDDVPVGNHSPKAPRKPRSPYKVLADAPALODDFYLNLDVMSHN 180
 QY 181 VLAAGLNCYVLLWNAACSSKVTKLCDLGVDDCVCSYGMORGHLAGVNNKVOIWDAR 240
 DB 181 VLAAGLNCYVLLWNAACSSKVTKLCDLGVDDCVCSYGMORGHLAGVNNKVOIWDAR 240
 QY 241 CKKIRSMEGHRLRVGLAWSSSLSSGGDKNITQDRIETQEDFVSKLSGHSKSEVCLKW 300
 DB 241 CKKIRSMEGHRLRVGLAWSSSLSSGGDKNITQDRIETQEDFVSKLSGHSKSEVCLKW 300
 QY 301 SYDNRELASGGDNKLFVWNOHSTQPVLYKCEHTAAVKAIAWSPHLGLASGGTADRC 360
 DB 301 SYDNRELASGGDNKLFVWNOHSTQPVLYKCEHTAAVKAIAWSPHLGLASGGTADRC 360
 QY 361 IREWNTTNSHLSGMDTSGVNCNLWMSKNVNLVSTHGYSONOIIYWRYPYMSKLTATL 420
 DB 361 IREWNTTNSHLSGMDTSGVNCNLWMSKNVNLVSTHGYSONOIIYWRYPYMSKLTATL 420
 QY 421 HTYRVLYLAISPDDGQITVTGAGDETLRFVWVFPSPKSQNTSEIGALSIGRTTIR 475
 DB 421 HTYRVLYLAISPDDGQITVTGAGDETLRFVWVFPSPKSQNTSEIGALSIGRTTIR 475

RESULT 3
 O82740 PRELIMINARY: PRT: 444 AA.
 AC O82740;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
 DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE PUTATIVE FIZZY-RELATED PROTEIN.
 GN FTH19.90 OR ATAG22910.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RN [1]
 NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RC Peters S.A., van Staveren M., Dirkse W., Stiekema W., Bancroft I.,
 RA Meves H.W., Mayer K.F.X., Scheller C., Bevan M.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AL031018; CA119806.1;
 DR EMBL: AL161558; CAB79246.1;
 DR InterPro: IPR000002; Fizzzy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.7.
 DR PRINTS: PRO0320; GPROTEINBRPT.
 DR PRODOM: PD004563; Fizzzy.1.
 DR SMART: SM00320; WD40.5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.

Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Gaps
1 MDGTGNRPPTSTVRDNDSP-----PEPSPELRVSRMKNHNTSPSKRTYSDRFTS	71.4%	350	1825	DB 10:	482	30	7
1 MEDESTTEPKKKSOSQNLPPSMNRPTVSLES-RINLIDSNHYSBSPKPYSDRFTS	72.9%	46	149	Indels			
1 RSASKRALFDI--NTPP--EGRDDSSAYTTLRLTALFGPOVAGPVPTEKTDSPMTLNP							
57 RSGSNALPDLWSSSPKKNKCKEDGAGSYASLKLTLF-----GVPEPKSDVYNGESPS							
59 RSGSNALPDLWSSSPKKNKCKEDGAGSYASLKLTLF-----GVPEPKSDVYNGESPS							
113 RNFRRKKTETROSMSHLSPEMDDDFVGVMHSPVKAPKRPVPSPKVLDADALODDEFLN							
114 GNIFREFTEFORSNLNTPPF-DSDVGVSGVSPSPKSPKRLKLSFKVLDADALODDEFLN							
173 LVDMSNNVAVLAVGLNCVYLLANACSSK-----VTKLCLDGLDNDCCVSGMA							
173 LVDMSNNVAVLAVGLNCVYLLANACSSK-----VTKLCLDGLDNDCCVSGMA							
219 ORGTHLANGTNGKVOJIMDAARCKTIRSEHGRIYVGLAWSSSLSSGGROKNIYORI							
233 LRKTHLALTSGGVYQIMDVYLRCKNIRTMGRLVGLAWSSSVLSSGRDSSIQRDI							
279 RTOEDFVSKLSHKEVCGELKMSYDNKELASGGDNKLFVYMHOSHSTOPLYKYCEHTAANK							
293 RTOEDFVSKLSHKEVCGELKMSYDNKELASGGDNKLFVYMHOSHSTOPLYKYCEHTAANK							
339 AIAMSPHLHGLLASGGGTADRCIRFWMTTNTNSHLSGMDTGSOVYCNLYMSKNVNLVSTHG							
353 AIAMSPHLHGLLASGGGTADRCIRFWMTTNTNTNLNCVDTISOVYCNLYMSKNVNLVSTHG							
399 YSONDIIYWRPMTSKLATTGHTYRVLVYLASPDGOTIYTGAGDETLRFVWVPSPSKQ							
413 YSONDIIYWRPMTSKLATTGHTYRVLVYLASPDGOTIYTGAGDETLRFVWVPSPSKQ							

RA Sato S., Kotani H., Nakamura Y., Kaneke T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.,
 RT Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT clones."
 CC DNA Res. 4:215-230(1997).
 DR -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AB005230. BAB1112.1.
 DR InterPro: IPR000002; F1zzy.
 DR Pfam: PF001680; WD40.
 DR ProDom: PD004563; F1zzy; 1.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR PROSITE: PS50082; WD_REPEATS_2; 4.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat
 SQ SEQUENCE 472 AA; 51330 MW; A9C6BAD79B5DAFB3 CRC64;

Query Match
 Best Local Similarity 56.7%; Score 1448; DB 10; Length 472;
 Matches 277; Conservative 60; Mismatches 90; Indels 22; Gaps 7;
 Oy 23 EPSPESLRHVRMINSHHTSPSRITTYSDRFIPSRASAKPALFDI-----NTPTEGRDSDS 77
 Db 25 ETSSSSFRGIISSL-----SSPSKSCSDRFIPCRSSRLHAFDLODEPTTPV--KEGG 76
 Oy 78 SAAVTLRLTALFGDVAGP-VTPE--KTDSPTMTLPNNITRYTETROSMSLSPPMD 134
 Db 77 NEAYSRLKSELFGSDPRLSLSPAGGQSGASSPSCPTNMLRKTRETSNPS--SPFSP 135
 Oy 135 DDFVPGVNH-----PVKARVPRSPRYVLAAPLADDFYLVDMSSHNVLAVGLGC 189
 Db 136 SILGNDGSHSSDPPPKPRKPKRKHVLDAPSLDDDFYLVDMSSHNVLAVGLGC 189
 Oy 190 VYLMNACSSKYTKICDGLGVDVCVGAAGTGLHAGTNNGVOTMDARCKTIRSMEG 249
 Db 196 VYLMNACSSKYTKICDGLGVDVCVGAAGTGLHAGTNNGVOTMDARCKTIRSMEG 249
 Oy 250 HRLVAGLAWSSSLSSGGRKNITORDIRTOEDFVSKISGKSEYCGILKMSYDRELAS 309
 Db 256 HQTRTGVLAWSRILSSGSRDRNLTQDIDVQSDPYFVSKISGKSEYCGILKMSYDRELAS 309
 Oy 310 GGNDDKLFVWNOHSTQPVLYKCEHTAAVKAIAWSPHLLGGLLAGGGTADRCIRFMMNTTN 369
 Db 316 GGNDDKLFVWNOHSTQPVLYKCEHTAAVKAIAWSPHLLGGLLAGGGTADRCIRFMMNTTN 369
 Oy 370 SHLSMDTGSQVCLVWKNVNEIYTHGYSQNOIIVRYPTMSKATLGLHTYRVLYLA 429
 Db 376 NQLSNIDTGSQVCLVWKNVNEIYTHGYSQNOIIMLWKPMSKATLGLHTYRVLYLA 429
 Oy 430 ISPDGQITVAGDETLRWNVFPSPKSO 458
 Db 436 ISPDGQITVAGDETLRWNVFPSPKSO 458

RESULT 6
 ID 075869 PRELIMINARY; PRT; 493 AA.
 AC 075869.
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DE R3374-2001 (Tremblrel. 19, last annotation update)
 GN R3374_1 (HYPOTHEICAL 54.8 KDA PROTEIN).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1] Taxid-9606;
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Vismathan V.,

RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillewagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 RA Danganan L., Eiler A., Christensen M., Georgescu A., Ayala J., Liu S.,
 RA Actix C., Andreise T., Frankheim M., Amco-Keller G., Coefield J.,
 RA Duarte S., Lucase S., Bruce R., Thomas P., Quan G., Krommiller B.,
 RA Arelano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RA Olsen A.S., Carraro A.V.;
 RT "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
 RT serine protease gene cluster";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Jin D.-Y., Jeang K.-T.;
 RT "Characterization of human fizzy-related protein";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN, NEUROBLASTOMA;
 RA Strausberg R.;
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AC005787; AAC62835.1;
 DR EMBL: AF085786; AAC62836.1;
 DR EMBL: AF083810; AAC52030.1;
 DR EMBL: BC013413; AAH13413.1;
 DR InterPro: IPR000002; F1zzy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR PRINTS: PR00320; GPROTEINRPT.
 DR ProDom: PD004563; F1zzy; 1.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
 DR PROSITE: PS50082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat; Hypothetical protein.
 SQ SEQUENCE 493 AA; 54765 MW; 766A947EB8BA6D5C CRC64;

Query Match
 Best Local Similarity 50.6%; Score 1294; DB 4; Length 493;
 Matches 269; Conservative 66; Mismatches 106; Indels 54; Gaps 14;
 Oy 6 NRNPSTSVDRNSPPSPESLRHVRMINSHHTSPSRITTYSDRFIPSRASAKPALFDI 64
 Db 17 NENTMPRTVETMRKRLTPASSPV-----SPSK--HGDRIPSRAGAMSVN 60
 Oy 65 -FDIN-----TPEGR--DDSS-----AYTLRLTALFG--PDVAGVPTPEKTDSPS 107
 Db 61 FHRINENKSPSONRRKAKDATSDNGDGLATLSALKKELLGAGIEKVDOPQEDRLDPS 120
 Oy 108 WTLNRRNIFRYKTKETROSK-----HSLSPMDDFVGVNHPVAKRKYRSPYK 158
 Db 121 -TPEKGLFTYSLSTRKSSPDGDNDVSEYSLSPVSNKS--QKLSRPKRPTRKISKIPFK 177
 Oy 159 VLDAPALDDDFYLVDMSSHNVLAVGLGNCVYIMNACSSKYTKICDGLVD--DCVCGVM 217
 Db 178 VLDAPALDDDFYLVDMSSHNVLAVGLGNCVYIMNACSSKYTKICDGLVD--DCVCGVM 217
 Oy 218 AORGHILVAGTNGSKVOIIMDAARCKIRSMEGHRLRVGALAMSSSLSSGGRKNITVND 277
 Db 238 SERGMLVAVGHNKFEVOIIMDAAGKRLMLBGRHARKVLAALMNAEQLSSGRBRLMLDND 297
 Oy 278 IRT-----DEFEVSKLSGKHSKSEYGLKSYDNNELASGGNDKLFVWNOHSTQPVLYKCEHT 334
 Db 298 IRTPELQSE--RLQGHQEVCGGLMWSHDHOLLASGGNDKLLVWNNHSLSPVQDYTEHL 355
 Oy 335 AAVKAIAMSPHLLGGLLAGGGTADRCIRFMMNTTNSHLSMDTGSQVCLVWKNVNEIY 394
 Db 356 AAVKAIAMSPHLLGGLLAGGGTADRCIRFMMNTLTGQPLDCLDITGSQVCLVWKNVNEIY 415
 Oy 395 STHGYSQNOIIVWRYPTMSKATLGLHTYRVLYLAISPDGQITVAGDETLRWNVFPSPKSO 454
 Db 416 STHGYSQNOIIVWRYPTMSKATLGLHTYRVLYLAISPDGQITVAGDETLRWNVFPSPKSO 473

us-09-701-572-2.rsp

RESULT	7	
Q9ULH8		
ID Q9ULH8		
PRELIMINARY;		
PRT;		504 AA

AC	Q9UJH8:2000 (Tremblrel. 13, Created)
AD	01-MAY-2000 (Tremblrel. 13, last sequence update)
DT	01-MAY-2000 (Tremblrel. 13, last annotation update)
DT	01-DEC-2001 (Tremblrel. 19, last annotation update)
DE	KIAA1242 PROTEIN (FRAGMENT).
GN	KIAA1242.
OC	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=BRAIN;
RC	MEDLINE=20039619; PubMed=10574462;
RA	Nagase T., Ishikawa K., Kikuno R., Hirotsawa M., Nomura N., Ohara O.,
RT	"Prediction of the coding sequences of unidentified human genes. XV.
RT	The complete sequences in vitro."
RT	For large proteins in vitro."
RT	DNA Res. 6:337-345(1999).
RL	1. SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC	EMBL: AB033068; BAA6556.1; .
DR	InterPro: IPR001002; Fizzy.
DR	InterPro: IPR001680; WD40.
DR	Pfam: PF00400; WD40. 7.
DR	Pfam: PF00320; GPROTEINBPT.
DR	PRINTS: PD004563; Fizzy. 1.
DR	PRODom: WD40. 5.
DR	SMART: SM00320; WD_REPEATS_1; UNKNOWN_2.
DR	PROSITE: PS00678; WD_REPEATS_2; 3.
DR	PROSITE: PS50082; WD_REPEATS_REGION; 1.
DR	PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW	Repeat: WD repeat. 1
FT	NON_TER 1
SO	SEQUENCE 504 AA; 55931 MW; 2BEC3EC3D68B395C CRC64;

Query Match	50.6%	Score 1.29e+103	
Best Local Similarity	54.3%	Pred. No. 5.2e+103	
Matches	269	Conservative	66 Mismatches 106 Indels 54 Gaps 14
QY	6	NNRPPTSTVDNSPPPEPSPESLRVSMISNMHTSPSPKRTIISPRIFPSKASKFAL-	71
Db	28	NEHTMTRVTEMRITLTPASSVS-----SPSK--HGDFIPSRAGANMWSVN	107
QY	65	-FDIN-----TPREGR--DDSSS-----AYTTLRLTALFG--PDVAGFVPTPEKIDSPS	131
Db	72	FHRINENKSPSNKRKADATSDNGKDLGYSALLNELLGAGIEKVVDDPTEDRRLQPS	158
QY	108	MTLENRNIFRYKTKETROSM-----HSLSPFMDDFVPGVSNIPYKARAKVPSRPX	188
Db	132	-TPREKGLFYFSLSTRKSSPDDGNDVSYLSLPSYKNS--QKLRSPKPRTKISKLPFK	217
QY	159	VIDARALDDDFYLLVYMSSHNVLAVGLAGNCVYUWMNCSKRYTKLDLGDV-DCYCSVGM	248
Db	189	VIDAPELDDDFLLVYMSSNLAVSLGTCVYLMASCTQGYRLDCLSYEGDSTVSGW	277
QY	218	AORGTHAVGTNNKGVQINDARCKIKIRMEGHRLLRVALGALSSSLSSGGGRDKNIYORD	308
Db	249	SEKGNLLVAVGTTHKGVQIIMDAAGKRLSMLEGHTARVGLALMNEDELSSSRBMILLOD	334
QY	278	IRI-----QEDPVSKLSHKSEYCGKLMSTDNRELASGNGNKLFFVNOHSTQPVLYKCEHT	366
Db	309	IRTPPLQSE--RLQDGHREYCGKLMSTDIQQLLAGSGNKNKLLVMNHSLSLSPVOOYTEHL	394
QY	335	AAVRLAIVSPHLGLGLASGGGADRCIFNMVTTTNSHLSCHMTGSOYCNLVMSKNVVELY	

```
Dd      427  STHGTSQNLIVMWRKSYSLTQVAKLTGHSRYVLILAMS PDGEALVTYGADETLRFNNVF--    480
Qy      395  SHHGSQNOIIVRREPYTMSKLATLTGTYEVLIALSPDGOITVGAGDETLRFNNVPSS    454
Db      367  AAVKIAMSHQHGLASSGGTADRCLREFNELLTQSPLOCDIDISQYCNLAMSKHANLEYL    426

Dd      455  PRSQNTSEFIALSL    469
Qy      485  SKTRSTKEVSIVLP    499
```

RESULT 0
Q9UM11
ID Q9UM11
PRELIMINARY;
PRT; 496 AA

```

AC      09MAY-2000 (TREMBLrel_13, Created)
AD      01-MAY-2000 (TREMBLrel_13, Last sequence update)
DT      01-MAY-2000 (TREMBLrel_13, Last annotation update)
DT      01-DEC-2001 (TREMBLrel_19, Last annotation update)
DE      FZrl.
GN      Homo sapiens (Human).
OS      Homo sapiens: Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Eukaryota; Metazoa; Chordata; Catarrhini; Hominiidae; Homo.
CC      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX      NCBI_TaxID=9606;
RN      [1]
RS      SEQUENCE FROM N.A.
RA      Kotani S., Oyama T., Todokoro K.;
RP      Human homologue of Flizy-related protein." ;
RT      Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
RL      *Human (APR-1998) to the EMBL/Genbank/DBJ DOMAINS).
RR      -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC      EMBL: AB013462; BAA6954.1; "-
DR      InterPro: IPR001680; WD40.
DR      Pfam: PF004400; WD40; 7.
DR      Pfam: PD004563; Flizy; 1.
DR      ProDom: SMO0320; WD40; 5.
DR      SMART: SMO0320; WD REPEATS_1; UNKNOWN_2.
DR      PROSITE: PS0082; WD REPEATS_2; 3.
DR      PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR      PROSITE: PS50294; WD_REPEATS_REGION; 1.
KM      Repeat: WD repeat.
SQ      Repeat: MD repeat.
        496 AA; 55188 MW; E555463E532DDC7 CRC64;
SEQUENCE

```

[illegible][illegible]

```

Db 356 AAVKAIAMSPHOGHLLASGGCTADRCIRFMWNTLTCGLOPIDTSSOVCLAMSKHANLEY 415
QY 395 STHGYSQNOIIVWRYPTMSKLTATLGHTRVYLALSPDGGOTITGAGDETLRFMNVFPS 454
Db 416 STHGYSQNOIIVWRYPTMSKLTATLGHTRVYLALSPDGGOTITGAGDETLRFMNVFPS 454
QY 455 PKSONTESE-IGALSL 469
Db 476 STRTKYKMSVSVLNL 491

RESULT 9
ID 09RKS PRELIMINARY; PRT: 493 AA.
AC 09RKS:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE FIZZY-RELATED PROTEIN (FZRI PROTEIN).
GN FZRI OR FYR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_Taxid=10090.
RP SEQUENCE FROM N.A.
RA Jin D.-Y., Jeang K.-T.
RT "Characterization of mouse fizzy-related protein."
RN MIMT000002 (AUG-1998) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RA Strausberg R.
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF083809; AAD52029.1;
DR MGD: BC006616; AAH06616.1;
DR InterPro: IPR000002; Fizzy.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR ProDom: PD004563; Fizzy; 1.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS50082; WD_REPEATS_2; 3.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SO SEQUENCE 493 AA; 54688 MM; B5DC6653D74D9A5 CRC64;

```

Query Match 50.3%; Score 1286; DB 11; Length 493;
 Best Local Similarity 53.9%; Pred. No. 2,5e-102;
 Matches 267; Conservative 69; Mismatches 105; Indels 54; Gaps 14;

```

Db 6 NNNPPTSTVDNPPPEPESLRLVSRMINSNHYTSPRTIYSDRFPSRSKAFAL- 64
QY 17 NENYPCVSEMRRTLTPTANSVPS-----SPSK--HGDRFIPRAGANMSVNL 60
Db 65 -FDIN---TPTEGR--DSSS-----AYTLRLTALFG---PDVAGPYTPKTDSPS 107
QY 61 FHRINENKSPSONRKAKDATSDNGKGLAYSLAKNELLGAGIEKVOPOEDRRLPS 120
Db 108 MTLPLNRIFRYKTTRO------HSLSPMDDFVPGVNHSPYKAPRRVPSYK 158
QY 121 -TPEHNGLETTYSLSKRSPPDGDNDVPSYLSPSNKS--OKLSPRKRPTKRTSKIPFK 177
Db 159 VLDAPALODDFYLNLYDMSSHNLYAVGLGNCVYLMAGSSKVTIKLCLDGLVD-DCVCSVGW 217
QY 178 VLDAPALODDFYLNLYDMSSHNLYAVGLGNCVYLMAGSSKVTIKLCLDGLVD-DCVCSVGW 217
Db 218 AORGTILAVGNGNGKQVIMDARCKIRMECHRLRVGALAMSSSLSSGGRKNTYOD 277
QY 238 SERGNLVAVTTHGFGVOIMDAAGKLEMLGHTARVGAALAMNADOLSSGSDRMLTORD 297

```

```

QY 278 IRT---QEDFVSKLSGKSEVCLKWSYDNRELASGNDKLFYMNHSTOPLKYCEHT 334
Db 298 IRTPELOSE--RLQGRDEVCGLKWSNDHOLLASGNDKLLVYMHSSLSIPVOQYTEHL 355
QY 335 AAVKAIAMSPHILHLLASGGCTADRCIRFMWNTLTCGLOPIDTSSOVCLAMSKHANLEY 415
Db 356 AAVKAIAMSPHOGHLLASGGCTADRCIRFMWNTLTCGLOPIDTSSOVCLAMSKHANLEY 415
QY 395 STHGYSQNOIIVWRYPTMSKLTATLGHTRVYLALSPDGGOTITGAGDETLRFMNVFPS 454
Db 416 STHGYSQNOIIVWRYPTMSKLTATLGHTRVYLALSPDGGOTITGAGDETLRFMNVFPS 454
QY 455 PKSONTESE-IGALSL 469
Db 474 SKTRTKESVSVLNL 488

```

```

RESULT 10
ID 042585 PRELIMINARY; PRT: 493 AA.
AC 042585:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
GN FIZZY-RELATED PROTEIN.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
RN NCBI_Taxid=8355;
RP SEQUENCE FROM N.A.
RA TISSUE-OVARY;
RC MEDLINE-97433078; PubMed-9288747;
RA Sigris S.J., Lehner C.F.
RT "Xenopus laevis fizzy-related down-regulates mitotic cyclins and is
RT required for cell proliferation arrest and entry into endocycles."
RC Cell 90:671-681(1997).
DR EMBL: Y14163; CA474576.1;
DR InterPro: IPR000002; Fizzy.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR ProDom: PD004563; FIZZYINBPT.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS50082; WD_REPEATS_2; 3.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SO SEQUENCE 493 AA; 54699 MM; BA19741AD3C6B6A7 CRC64;

```

Query Match 50.3%; Score 1285; DB 13; Length 493;
 Best Local Similarity 53.7%; Pred. No. 3e-102;
 Matches 266; Conservative 70; Mismatches 105; Indels 54; Gaps 14;

```

Db 6 NNNPPTSTVDNPPPEPESLRLVSRMINSNHYTSPRTIYSDRFPSRSKAFAL- 64
QY 17 NENYPCVSEMRRTLTPTANSVPS-----SPSK--HGDRFIPRAGANMSVNL 60
Db 65 -FDIN---TPTEGR--DSSS-----AYTLRLTALFG---PDVAGPYTPKTDSPS 107
QY 61 FHRINENKSPSONRKAKDATSDNGKGLAYSLAKNELLGAGIEKVOPOEDRRLPS 120
Db 108 MTLPLNRIFRYKTTRO------HSLSPMDDFVPGVNHSPYKAPRRVPSYK 158
QY 121 -TPEHNGLETTYSLSKRSPPDGDNDVPSYLSPSNKS--OKLSPRKRPTKRTSKIPFK 177
Db 159 VLDAPALODDFYLNLYDMSSHNLYAVGLGNCVYLMAGSSKVTIKLCLDGLVD-DCVCSVGW 217

```


RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AEO03431; AAF45973.1; -
 DR Flybase: FBgn0003200; rap.
 DR InterPro: IPR000002; F1zy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR Prodom: PD004563; F1zy; 1.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS00678; WD_REPEATS_1; 2.
 DR PROSITE: PS0082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 478 AA; 53507 MW; 0E6098740897CF76 CRC64;

Query Match 49.9%; Score 1276; DB 5; Length 478;
 Best Local Similarity 54.9%; Pred. No. 1.7e-101;
 Matches 260; Conservative 62; Mismatches 114; Indels 38; Gaps 11;

QY 22 PEPESLRHVSRLNINS--NHYSPTSTYSDRFPSPS-----ASKALFDIN----- 68
 DB 4 PEYKRLKHYSPVARNLNFNFESSSTPTSLDRFLPCAYNNMOTJNA--SINKSDNSP 61
 QY 69 -TPEGRD-----DSSAYTLRLTALFGRPV-----AGPYTEKTPSPMTLPNENITR 117
 DB 62 QTSKRDGCGEATROSLAYSLKLNELLGSAIDVKTAGEENENAYTRA--AKSLK 118
 QY 118 YKTEQDSM-----HSLSPFMDDFVPCVNSPVKAPRKVRSPRYVLAAPLLODFEYL 172
 DB 119 YQSPTRQDYNGCEPTSLSPVSAKS--OKLLRSPRKATRKISRLPFVLDAPLDDFEYL 176
 QY 173 LVDNSHNVLAVALGNCVYLLNACSSKVTLCGLVD--DCCVSCVMAQRGTHLAVGTNG 231
 DB 177 LVDMSQNVLAVALGSCVYLLNACSSKVTLCGLVD--DCCVSCVMAQRGTHLAVGTNG 231
 QY 232 KVALWDARCKKIRSMGHRRLRGALAMSSSLSSGGRDKNTYORDIF--OEDVSKLSG 290
 DB 237 YVFWVDAVANKQIKKLMGSHARVALAMNSDLSGSRDMLIQDRITPOLOSERLAG 296
 QY 291 HKSPVCLAKSYNRELASGNDKLFVNQHSOTQPVLYKCEHRAVKAIAVSPHLLGL 350
 DB 297 HRQEVGCLKWSPDNOYLASGNDRLYVWNGHSAVPOSTTEHMAAVKAIAMSPHHGL 356
 QY 351 ASSGGTADRCIRFVNNTTNSHLSGMDTSGOVCLVMSKNVNELVSTHGSQNOIIVMRP 410
 DB 357 ASSGGTADRCIRFVNNTTNSHLSGMDTSGOVCLVMSKNVNELVSTHGSQNOIIVMRP 410
 QY 411 TMSKLTATLGHYRVLVLAISPQDQITVAGDETLFVNVPSPSQMTESEI 464
 DB 417 SLTQVAKLTGHSYRVLVLAISPQDQITVAGDETLFVNVPSPSQMTESEI 464
 RESULT 13
 ID 090196 PRELIMINARY; PRT; 493 AA.
 AC 090196;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE 01-DEC-2001 (Tremblrel. 13, Last annotation update)
 GN F1ZY-RELATED PROTEIN HOMOLOG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCB1_TaxID=9606;
 RX [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Suo T., Sava H.;
 RT "Identification of a human homolog of the *Drosophila* fuzzy-related
 protein."
 RL Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AF080397; AAF20266.1; -
 DR InterPro: IPR000002; F1zy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR Prodom: PD004563; F1zy; 1.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN.2.
 DR PROSITE: PS0082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 493 AA; 54905 MW; 9DE5DA138C19383C CRC64;

Query Match 49.9%; Score 1274; DB 4; Length 493;
 Best Local Similarity 53.7%; Pred. No. 2.7e-101;
 Matches 266; Conservative 67; Mismatches 108; Indels 54; Gaps 14;

QY 6 NRPPTSTYRDNSSPPSPESLRHVSRLNINSNHYSPTSTYSDRFPSPSASKAL- 64
 DB 17 NENTMRYTEMKRTITLPASSPVS-----SPSK--HGRTFSPAGAMSVN 60
 QY 65 -FDIN-----TPEGR--DSSS-----AYTLRLTALFGR--PDVAGPYTEKTPSP 107
 DB 61 FRIENENKSPSONKAKADATSDNGKGLAYSLNELLGAGIERKVDQPTEDRRLOPS 120
 QY 108 MTLPRNIRFKYKTEKTSQSM-----HSLSPFMDDFVPCVNSPVKAPRKVRSPRY 138
 DB 121 -TPEKGLFTYSLTKSSPDCCNAXSPVSLSPVSAKS--OKLLRSPRKATRKISRLPF 177
 QY 159 VLDPALODDFEYLNLVDMSSHNVLAVALGNCVYLLNACSSKVTLCGLVD--DCCVSCV 217
 DB 178 VLDPALODDFEYLNLVDMSSHNVLAVALGNCVYLLNACSSKVTLCGLVD--DCCVSCV 217
 QY 218 AQRGTHLAVGTNGKQVQIDARCKKIRSMGHRRLRGALAMSSSLSSGGRDKNTYORD 277
 DB 238 SERGNLVAAGTHGKQVQIDARCKKIRSMGHRRLRGALAMSSSLSSGGRDKNTYORD 277
 QY 278 IRT--OEDVSKLSGSHSEVCLKWSYDRELASGNDKLFVNQHSOTQPVLYKCEH 334
 DB 298 IRTPLQSF--RLQSHQEVGCLKWSYDRELASGNDKLFVNQHSOTQPVLYKCEH 334
 QY 335 AAVKAIAMSPHHGLLASGGGTADRCIRFVNNTTNSHLSGMDTSGOVCLVMSKNVNELV 394
 DB 356 AAVKAIAMSPHHGLLASGGGTADRCIRFVNNTTNSHLSGMDTSGOVCLVMSKNVNELV 394
 QY 395 STHGSQNOIIVMRPPTMSKLTATLGHYRVLVLAISPQDQITVAGDETLFVNVPSP 415
 DB 416 STHGSQNOIIVMRPPTMSKLTATLGHYRVLVLAISPQDQITVAGDETLFVNVPSP 415
 QY 455 PKSONTESEIGALSL 469
 DB 474 SKTRSTKESVSVNL 488
 RESULT 14
 ID 094DF9 PRELIMINARY; PRT; 528 AA.
 AC 094DF9;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DE PUTATIVE CELL CYCLE SWITCH PROTEIN.
 GN P0518C01.29.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

